Potential Interference w/ Sheppard et al
RTN. 37

OM of: US-09-634-252A-2 to: A_Geneseq_0601:* out_format : pfs

Date: Aug 15, 2001 9:55 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09634252/runat_14082001_142627_9794/app_query.fasta_1.2598
-DB=A_Geneseq_0601 -QFMT=fastan -SUFFIX=n2p.rag -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
[redacted]LOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
[redacted]NS=human40.cdi -LIST=90 -DOCALIGN=200 -THR_SCORE=pct
[redacted]_MAX=100 -THR_MIN=0 -ALIGN=30 -MODE=LOCAL -OUTFMT=pfs
[redacted]ORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09634252_@CGN1_1_82 -NCPU=6 -ICPU=3 -LONGLOG -NC_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-634-252A-2
Query length: 2499
Database: A_Geneseq_0601:*
Database sequences: 412676
Database length: 60623988
Search time (sec): 81.930000

Cerretti

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	!	Documentation	..			
/SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY30207	+	4553.00	6360.69	0.0	832	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY25120	+	4553.00	6360.69	0.0	832	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY54457	+	3621.00	5055.37	1.5e-274	696	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY30208	+	1985.00	2758.29	1.1e-146	867	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY25119	+	1983.50	2756.75	1.4e-146	823	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY25118	+	1983.50	2756.29	1.4e-146	859	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:AAR75352	+	1978.00	2749.77	3.6e-146	769	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:AAR67759	+	1862.50	2589.26	3.6e-137	670	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:AAR67758	+	1297.00	1798.73	5.0e-93	524	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:AAR67757	+	1293.00	1793.89	1.0e-92	488	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB50942	+	1007.50	1388.40	2.4e-70	778	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB50935	+	1007.50	1386.61	2.6e-70	918	!					
[redacted]DS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW25716	+	975.00	1341.00	9.0e-68	920	!					
[redacted]S8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW25715	+	966.00	1328.58	4.5e-67	903	!					
[redacted]S8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY41695	+	965.50	1330.10	4.5e-67	735	!					
[redacted]DS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB44251	+	965.50	1330.10	4.5e-67	735	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW25719	+	951.00	1310.51	6.0e-66	686	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB47106	+	897.00	1232.94	1.1e-61	812	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW25717	+	892.00	1225.50	2.7e-61	845	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB47105	+	871.00	1196.61	1.1e-59	802	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW01825	+	829.00	1138.41	2.1e-56	751	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAW90855	+	812.50	1114.93	4.2e-55	775	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAW90865	+	810.50	1112.12	6.0e-55	775	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW44120	+	774.00	1062.66	4.0e-52	660	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW73009	+	772.50	1061.36	5.1e-52	613	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW73013	+	766.50	1052.80	1.5e-51	621	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:AAR99801	+	765.50	1049.97	1.9e-51	709	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY17413	+	765.00	1049.01	2.1e-51	726	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY28655	+	760.00	1042.05	5.1e-51	722	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW73007	+	751.50	1031.76	2.2e-50	621	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW73011	+	749.00	1028.77	3.4e-50	592	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY03223	+	749.00	1026.57	3.7e-50	726	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB20251	+	745.00	1019.59	8.0e-50	824	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:AAR87037	+	739.00	1012.42	2.3e-49	734	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB61896	+	736.00	1010.11	3.6e-49	616	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB07742	+	720.00	984.98	7.1e-48	790	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB07743	+	719.00	983.70	8.4e-48	781	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB07705	+	719.00	983.57	8.5e-48	790	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB07739	+	717.00	981.10	1.2e-47	766	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB07740	+	717.00	980.81	1.2e-47	787	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB07741	+	717.00	980.36	1.2e-47	820	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:AAR87034	+	711.00	974.45	3.3e-47	651	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:AAR87036	+	691.00	945.08	1.3e-45	735	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY03224	+	689.50	946.27	1.5e-45	542	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW73010	+	649.00	889.89	2.1e-42	521	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW25722	+	635.50	873.98	2.2e-41	394	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAW90851	+	624.00	854.44	1.9e-40	540	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAW90864	+	624.00	854.44	1.9e-40	540	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW73012	+	611.00	837.89	1.9e-39	462	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW25723	+	601.00	827.77	9.9e-39	322	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW14772	+	588.00	804.15	1.2e-37	530	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG04037	+	573.00	798.82	1.1e-36	124	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB07738	+	548.50	749.39	1.5e-34	499	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB64744	+	525.50	720.22	8.3e-33	375	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB64745	+	522.00	715.55	1.5e-32	367	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB50937	+	522.00	713.71	1.7e-32	435	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB64681	+	522.00	713.71	1.7e-32	435	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:AAR87035	+	517.00	706.16	4.1e-32	457	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY17414	+	514.00	700.25	7.6e-32	535	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW73008	+	486.50	663.82	9.9e-30	439	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY79413	+	476.00	648.06	6.8e-29	483	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT:AAR42867	+	465.50	633.40	4.5e-28	480	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT:AAR52634	+	461.00	630.41	8.9e-28	353	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB51995	+	422.50	579.63	8.1e-25	262	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB51993	+	414.50	565.78	3.7e-24	334	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB51994	+	411.50	561.58	6.4e-24	334	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:AAP90459	+	401.50	543.67	4.5e-23	478	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:AAP90461	+	401.50	543.67	4.5e-23	478	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:AAR04083	+	401.50	543.67	4.5e-23	478	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:AAP90458	+	397.50	539.86	8.6e-23	405	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:AAR04082	+	397.50	539.86	8.6e-23	405	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW28509	+	397.50	538.25	9.1e-23	470	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW75769	+	397.50	538.25	9.1e-23	470	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:AAP90462	+	392.50	531.46	2.2e-22	460	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:AAP90457	+	383.50	520.06	1.1e-21	411	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT:AAR04081	+	383.50	520.06	1.1e-21	411	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT:AAP90463	+	383.50	518.89	1.1e-21	458	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW40991	+	362.00	482.38	6.7e-20	824	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB51996	+	359.00	490.40	7.4e-20	266	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW08410	+	353.50	472.52	2.9e-19	681	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW72743	+	353.50	472.52	2.9e-19	681	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW08406	+	353.50	470.68	3.1e-19	807	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW72737	+	353.50	470.68	3.1e-19	807	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW25721	+	345.00	478.43	7.0e-19	131	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW56134	+	332.50	442.03	1.3e-17	749	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY16776	+	331.50	440.65	1.6e-17	748	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW70457	+	330.50	439.24	1.9e-17	748	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB72283	+	318.00	418.60	2.0e-16	997	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB72301	+	318.00	411.43	2.6e-16	1934	!					
/SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW35293	+	317.00	419.59	2.2e-16	799	!					

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY30207

seq_documentation_block:
ID AAY30207 standard; Protein; 832 AA.
XX
AC AAY30207;
XX
DT 29-OCT-1999 (first entry)
XX
DE Amino acid sequence of the human SVPH3-17 protein.
XX
KW SVPH3-17; metalloproteinase-disintegrin; human chromosome 2;
KW human chromosome 7; malignant hyperthermia susceptibility;
KW Zellweger syndrome; neonatal adrenoleukodystrophy;
KW infantile Refsum disease; progressive familial intrahepatic colchestatis;
KW mucopolysaccharidosis VII; split hand/foot malformation;
KW arrhythymogenic right ventricular displasia-4; Coppock-like cataract;
KW insulin dependent diabetes mellitus-12; lamellar type ichthyosis;
KW transient neonatal myasthenia gravis; congenital aculeiform cataract;
KW juvenile amyotrophic lateral sclerosis;
KW familial paroxysmal choreoathetosis;
KW Finnish lethal neonatal metabolic syndrome.

XX
OS Homo sapiens.
XX
PN WO9941388-A2.
XX
PD 19-AUG-1999.
XX
PF 11-FEB-1999; 99WO-US03016.
XX
PR 11-FEB-1998; 98US-0074310.
XX
PA (IMMV) IMMUNEX CORP.
XX
PI Cerretti DP;
XX
DR WPI; 1999-527371/44.
DR N-PSDB; AAZ10207.
XX
PT DNA encoding the SVPH3-13 and SVPH3-17 proteins for detecting
PT disease corresponding to chromosome 7, e.g. Zellweger syndrome
XX
[redacted] Claim 1; Page 8-9; 82pp; English.
XX
CC The present sequence represents a protein designated SVPH3-17, which
CC is a member of the metalloproteinase-disintegrin family. The
CC specification also describes SVPH3-13 proteins. Both sequences can
CC be used to identify human chromosome 2 or 7, and to map genes on these
CC two chromosomes, and also to identify genes associated with certain
CC diseases, syndromes, or other human conditions associated with human
CC chromosome 2 or 7. The disease that correspond to chromosome 7 include
CC malignant hyperthermia susceptibility, Zellweger syndrome, neonatal
CC adrenoleukodystrophy, infantile Refsum disease, progressive familial
CC intrahepatic colchestatis, mucopolysaccharidosis VII, and split
CC hand/foot malformation. Diseases associated with chromosome 2 include
CC arrhythymogenic right ventricular displasia-4, insulin dependent diabetes
CC mellitus-12, transient neonatal myasthenia gravis, juvenile amyotrophic
CC lateral sclerosis, congenital aculeiform cataract, Coppock-like cataract,
CC lamellar type ichthyosis, familial paroxysmal choreoathetosis, and
CC Finnish lethal neonatal metabolic syndrome.
XX
SQ Sequence 832 AA;

alignment_scores:
Quality: 4553.00 Length: 832
Ratio: 5.472 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-634-252A-2 x AAY30207 ..

Align seg 1/1 to: AAY30207 from: 1 to: 832

```
  1 ATGAAGCCGCCCGGCAGCAGCTCGCGGCAGCCGCCCCTGGCGGGCTGCAG 50
    |||||||||||||||||||||||||||||||||||||||||||||||||||
  1 MetLysProProGlySerSerSerArgGlnProProLeuAlaGlyCysSe 17

 51 CCTTGCCGGCGCTTCCTGCGGCCCCCAACGCGGCCCCGCCGGCTCGGTGC 100
    |||||||||||||||||||||||||||||||||||||||||||||||||||
 17 rLeuAlaGlyAlaSerCysGlyProGlnArgGlyProAlaGlySerValP 34

101 CTGCCAGCGCCCCGGCCCGCACGCCGCCCTGCCGCCTGCTTCTCGTCCTT 150
    |||||||||||||||||||||||||||||||||||||||||||||||||||
 34 roAlaSerAlaProAlaArgThrProProCysArgLeuLeuLeuValLeu 50

151 CTCCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCGCGCCTGGGGGGC 200
    |||||||||||||||||||||||||||||||||||||||||||||||||||
 51 LeuLeuLeuProProLeuAlaAlaSerSerArgProArgAlaTrpGlyAl 67

201 TGCTGCGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAAAAAAATTTGG 250
    |||||||||||||||||||||||||||||||||||||||||||||||||||
 67 aAlaAlaProSerAlaProHisTrpAsnGluThrAlaGluLysAsnLeuG 84

251 GAGTCCTGGCAGATGAAGACAATACATTGCAACAGAATAGCAGCAGTAAT 300
    |||||||||||||||||||||||||||||||||||||||||||||||||||
 84 lyValLeuAlaAspGluAspAsnThrLeuGlnGlnAsnSerSerSerAsn 100

301 ATCAGTTACAGCAATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACT 350
    |||||||||||||||||||||||||||||||||||||||||||||||||||
101 IleSerTyrSerAsnAlaMetGlnLysGluIleThrLeuProSerArgLe 117

351 CATATATTACATCAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACA 400
    |||||||||||||||||||||||||||||||||||||||||||||||||||
117 uIleTyrTyrIleAsnGlnAspSerGluSerProTyrHisValLeuAspT 134

401 CAAAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAG 450
    |||||||||||||||||||||||||||||||||||||||||||||||||||
134 hrLysAlaArgHisGlnGlnLysHisAsnLysAlaValHisLeuAlaGln 150

451 GCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCAT 500
    |||||||||||||||||||||||||||||||||||||||||||||||||||
151 AlaSerPheGlnIleGluAlaPheGlySerLysPheIleLeuAspLeuIl 167

501 ACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAA 550
    |||||||||||||||||||||||||||||||||||||||||||||||||||
167 eLeuAsnAsnGlyLeuLeuSerSerAspTyrValGluIleHisTyrGluA 184

551 ATGGGAAACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGA 600
    |||||||||||||||||||||||||||||||||||||||||||||||||||
184 snGlyLysProGlnTyrSerLysGlyGlyGluHisCysTyrTyrHisGly 200

601 AGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGG 650
    |||||||||||||||||||||||||||||||||||||||||||||||||||
201 SerIleArgGlyValLysAspSerLysValAlaLeuSerThrCysAsnGl 217

651 ACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCAC 700
    |||||||||||||||||||||||||||||||||||||||||||||||||||
217 yLeuHisGlyMetPheGluAspAspThrPheValTyrMetIleGluProL 234

701 TAGAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCACATATAATCCAG 750
    |||||||||||||||||||||||||||||||||||||||||||||||||||
234 euGluLeuValHisAspGluLysSerThrGlyArgProHisIleIleGln 250

751 AAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGA 800
    |||||||||||||||||||||||||||||||||||||||||||||||||||
251 LysThrLeuAlaGlyGlnTyrSerLysGlnMetLysAsnLeuThrMetGl 267

801 AAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAA 850
    |||||||||||||||||||||||||||||||||||||||||||||||||||
267 uArgGlyAspGlnTrpProPheLeuSerGluLeuGlnTrpLeuLysArgA 284

851 GGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATAT 900
    |||||||||||||||||||||||||||||||||||||||||||||||||||
284 rgLysArgAlaValAsnProSerArgGlyIlePheGluGluMetLysTyr 300

901 TTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTC 950
    |||||||||||||||||||||||||||||||||||||||||||||||||||
301 LeuGluLeuMetIleValAsnAspHisLysThrTyrLysLysHisArgSe 317

951 TTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGG 1000
    |||||||||||||||||||||||||||||||||||||||||||||||||||
317 rSerHisAlaHisThrAsnAsnPheAlaLysSerValValAsnLeuValA 334

1001 ATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTA 1050
     |||||||||||||||||||||||||||||||||||||||||||||||||||
334 spSerIleTyrLysGluGlnLeuAsnThrArgValValLeuValAlaVal 350

1051 GAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCA 1100
     |||||||||||||||||||||||||||||||||||||||||||||||||||
351 GluThrTrpThrGluLysAspGlnIleAspIleThrThrAsnProValGl 367

1101 GATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATTAAGCAGCATGCTG 1150
     |||||||||||||||||||||||||||||||||||||||||||||||||||
367 nMetLeuHisGluPheSerLysTyrArgGlnArgIleLysGlnHisAlaA 384

1151 ATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGT 1200
```

```
 384 spAlaValHisLeuIleSerArgValThrPheHisTyrLysArgSerSer 400

1201 CTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAA 1250
 401 LeuSerTyrPheGlyGlyValCysSerArgThrArgGlyValGlyValAs 417

1251 TGAGTATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTGG 1300
 417 nGluTyrGlyLeuProMetAlaValAlaGlnValLeuSerGlnSerLeuA 434

1301 CTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATGT 1350
 434 laGlnAsnLeuGlyIleGlnTrpGluProSerSerArgLysProLysCys 450

 351 GACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTC 1400
 451 AspCysThrGluSerTrpGlyGlyCysIleMetGluGluThrGlyValSe 467

1401 CCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTT 1450
 467 rHisSerArgLysPheSerLysCysSerIleLeuGluTyrArgAspPheL 484

1451 TACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTACTT 1500
 484 euGlnArgGlyGlyGlyAlaCysLeuPheAsnArgProThrLysLeuPhe 500

1501 GAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGA 1550
 501 GluProThrGluCysGlyAsnGlyTyrValGluAlaGlyGluGluCysAs 517

1551 TTGTGGTTTTCATGTGGAATGCTATGGATTATGCTGTAAGAAATGTTCCC 1600
 517 pCysGlyPheHisValGluCysTyrGlyLeuCysCysLysLysCysSerL 534

1601 TCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCA 1650
 534 euSerAsnGlyAlaHisCysSerAspGlyProCysCysAsnAsnThrSer 550

1651 TGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTG 1700
 551 CysLeuPheGlnProArgGlyTyrGluCysArgAspAlaValAsnGluCy 567

1701 TGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATC 1750
 567 sAspIleThrGluTyrCysThrGlyAspSerGlyGlnCysProProAsnL 584

1751 TTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTAC 1800
 584 euHisLysGlnAspGlyTyrAlaCysAsnGlnAsnGlnGlyArgCysTyr 600

1801 AATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAAC 1850
 601 AsnGlyGluCysLysThrArgAspAsnGlnCysGlnTyrIleTrpGlyTh 617

1851 AAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAG 1900
 617 rLysAlaAlaGlySerAspLysPheCysTyrGluLysLeuAsnThrGluG 634

1901 GCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGC 1950
 634 lyThrGluLysGlyAsnCysGlyLysAspGlyAspArgTrpIleGlnCys 650

1951 AGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCG 2000
 651 SerLysHisAspValPheCysGlyPheLeuLeuCysThrAsnLeuThrAr 667

2001 AGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCT 2050
 667 gAlaProArgIleGlyGlnLeuGlnGlyGluIleIleProThrSerPheT 684

2051 ACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGAT 2100
```

Encodes ADAM27 possibly cleaved 489->569

dBEST Id No. 185722

R15308.1 Hillier et al. 1995 EST yf86d08.r1

```
 684 yrHisGlnGlyArgValIleAspCysSerGlyAlaHisValValLeuAsp 700

2101 GATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTC 2150
 701 AspAspThrAspValGlyTyrValGluAspGlyThrProCysGlyProSe 717

2151 TATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGA 2200
 717 rMetMetCysLeuAspArgLysCysLeuGlnIleGlnAlaLeuAsnMetS 734

2201 GCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTG 2250
 734 erSerCysProLeuAspSerLysGlyLysValCysSerGlyHisGlyVal 750

2251 TGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGA 2300
 751 CysSerAsnGluAlaThrCysIleCysAspPheThrTrpAlaGlyThrAs 767

2301 TTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAAGGATGAAG 2350
 767 pCysSerIleArgAspProValArgAsnLeuHisProProLysAspGluG 784

2351 GACCCAAGGGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGGT 2400
 784 lyProLysGlyProSerAlaThrAsnLeuIleIleGlySerIleAlaGly 800

2401 GCCATCCTGGTAGCAGCTATTGTCCTTGGGGGCACAGGCTGGGGATTTAA 2450
 801 AlaIleLeuValAlaAlaIleValLeuGlyGlyThrGlyTrpGlyPheLy 817

2451 AAATGTCAAGAAGAGAAGGTTCGATCCTACTCAGCAAGGCCCCATC 2496
 817 sAsnValLysLysArgArgPheAspProThrGlnGlnGlyProIle 832
```

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY25120

```
seq_documentation_block:
ID   AAY25120 standard; Protein; 832 AA.
XX
AC   AAY25120;
XX
DT   26-AUG-1999  (first entry)
XX
DE   Human MDC3 protein.
XX
KW   Metalloproteinase-like-disintegrin-like cysteine rich protein; human;
KW   MDC2-alpha; MDC2-beta; MDC3; medical; treatment; diagnosis.
XX
OS   Homo sapiens.
XX
PN   JP11155574-A.
XX
PD   15-JUN-1999.
XX
PF   01-DEC-1997;   97JP-0330020.
XX
PR   01-DEC-1997;   97JP-0330020.
XX
PA   (EISA ) EISAI CO LTD.
XX
DR   WPI; 1999-398071/34.
DR   N-PSDB; AAX78438.
XX
PT   New protein belonging to MDC gene family  - useful in medical
PT   treatment and diagnosis
XX
PS   Disclosure; Page 13-16; 17pp; Japanese.
XX
CC   This invention describes novel human MDC2-alpha, MDC2-beta and MDC3
CC   proteins and their encoding nucleic acids. The products of the invention
CC   are useful in medical treatment and diagnosis.
XX
SQ   Sequence   832 AA;
```

```
alignment_scores:
            Quality: 4553.00              Length:    832
              Ratio:  5.472                Gaps:        0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-634-252A-2 x AAY25120   ..

 Align seg 1/1  to: AAY25120  from: 1  to: 832

       1 ATGAAGCCGCCCGGCAGCAGCTCGCGGCAGCCGCCCCTGGCGGGCTGCAG 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetLysProProGlySerSerSerArgGlnProProLeuAlaGlyCysSe 17

      51 CCTTGCCGGCGCTTCCTGCGGCCCCCAACGCGGCCCCGCCGGCTCGGTGC 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 rLeuAlaGlyAlaSerCysGlyProGlnArgGlyProAlaGlySerValP 34

     101 CTGCCAGCGCCCCGGCCCGCACGCCGCCCTGCCGCCTGCTTCTCGTCCTT 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 roAlaSerAlaProAlaArgThrProProCysArgLeuLeuLeuValLeu 50

     151 CTCCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCGCGCCTGGGGGGC 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 LeuLeuLeuProProLeuAlaAlaSerSerArgProArgAlaTrpGlyAl 67

     201 TGCTGCGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAAAAAAATTTGG 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 aAlaAlaProSerAlaProHisTrpAsnGluThrAlaGluLysAsnLeuG 84

     251 GAGTCCTGGCAGATGAAGACAATACATTGCAACAGAATAGCAGCAGTAAT 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 lyValLeuAlaAspGluAspAsnThrLeuGlnGlnAsnSerSerSerAsn 100

     301 ATCAGTTACAGCAATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACT 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 IleSerTyrSerAsnAlaMetGlnLysGluIleThrLeuProSerArgLe 117

     351 CATATATTACATCAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACA 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 uIleTyrTyrIleAsnGlnAspSerGluSerProTyrHisValLeuAspT 134

     401 CAAAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAG 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 hrLysAlaArgHisGlnGlnLysHisAsnLysAlaValHisLeuAlaGln 150

     451 GCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCAT 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 AlaSerPheGlnIleGluAlaPheGlySerLysPheIleLeuAspLeuIl 167

     501 ACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAA 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 eLeuAsnAsnGlyLeuLeuSerSerAspTyrValGluIleHisTyrGluA 184

     551 ATGGGAAACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGA 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 snGlyLysProGlnTyrSerLysGlyGlyGluHisCysTyrTyrHisGly 200

     601 AGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGG 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 SerIleArgGlyValLysAspSerLysValAlaLeuSerThrCysAsnGl 217

     651 ACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCAC 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 yLeuHisGlyMetPheGluAspAspThrPheValTyrMetIleGluProL 234

     701 TAGAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCACATATAATCCAG 750
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 euGluLeuValHisAspGluLysSerThrGlyArgProHisIleIleGln 250

     751 AAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGA 800
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 LysThrLeuAlaGlyGlnTyrSerLysGlnMetLysAsnLeuThrMetGl 267

     801 AAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAA 850
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     267 uArgGlyAspGlnTrpProPheLeuSerGluLeuGlnTrpLeuLysArgA 284

     851 GGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATAT 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     284 rgLysArgAlaValAsnProSerArgGlyIlePheGluGluMetLysTyr 300

     901 TTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTC 950
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 LeuGluLeuMetIleValAsnAspHisLysThrTyrLysLysHisArgSe 317

     951 TTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGG 1000
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     317 rSerHisAlaHisThrAsnAsnPheAlaLysSerValValAsnLeuValA 334

    1001 ATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTA 1050
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     334 spSerIleTyrLysGluGlnLeuAsnThrArgValValLeuValAlaVal 350

    1051 GAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCA 1100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GluThrTrpThrGluLysAspGlnIleAspIleThrThrAsnProValGl 367

    1101 GATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATTAAGCAGCATGCTG 1150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     367 nMetLeuHisGluPheSerLysTyrArgGlnArgIleLysGlnHisAlaA 384

    1151 ATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGT 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     384 spAlaValHisLeuIleSerArgValThrPheHisTyrLysArgSerSer 400

    1201 CTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAA 1250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 LeuSerTyrPheGlyGlyValCysSerArgThrArgGlyValGlyValAs 417

    1251 TGAGTATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTGG 1300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     417 nGluTyrGlyLeuProMetAlaValAlaGlnValLeuSerGlnSerLeuA 434

    1301 CTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATGT 1350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     434 laGlnAsnLeuGlyIleGlnTrpGluProSerSerArgLysProLysCys 450

    1351 GACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTC 1400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     451 AspCysThrGluSerTrpGlyGlyCysIleMetGluGluThrGlyValSe 467

    1401 CCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTT 1450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     467 rHisSerArgLysPheSerLysCysSerIleLeuGluTyrArgAspPheL 484

    1451 TACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTT 1500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     484 euGlnArgGlyGlyGlyAlaCysLeuPheAsnArgProThrLysLeuPhe 500

    1501 GAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGA 1550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     501 GluProThrGluCysGlyAsnGlyTyrValGluAlaGlyGluGluCysAs 517

    1551 TTGTGGTTTTCATGTGGAATGCTATGGATTATGCTGTAAGAAATGTTCCC 1600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     517 pCysGlyPheHisValGluCysTyrGlyLeuCysCysLysLysCysSerL 534

    1601 TCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCA 1650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     534 euSerAsnGlyAlaHisCysSerAspGlyProCysCysAsnAsnThrSer 550

    1651 TGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTG 1700
         ||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
 551 CysLeuPheGlnProArgGlyTyrGluCysArgAspAlaValAsnGluCy 567

1701 TGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATC 1750
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 567 sAspIleThrGluTyrCysThrGlyAspSerGlyGlnCysProProAsnL 584

1751 TTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTAC 1800
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 584 euHisLysGlnAspGlyTyrAlaCysAsnGlnAsnGlnGlyArgCysTyr 600

1801 AATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAAC 1850
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 601 AsnGlyGluCysLysThrArgAspAsnGlnCysGlnTyrIleTrpGlyTh 617

1851 AAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAG 1900
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 617 rLysAlaAlaGlySerAspLysPheCysTyrGluLysLeuAsnThrGluG 634

1901 GCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGC 1950
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 634 lyThrGluLysGlyAsnCysGlyLysAspGlyAspArgTrpIleGlnCys 650

1951 AGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCG 2000
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 651 SerLysHisAspValPheCysGlyPheLeuLeuCysThrAsnLeuThrAr 667

2001 AGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCT 2050
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 667 gAlaProArgIleGlyGlnLeuGlnGlyGluIleIleProThrSerPheT 684

2051 ACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTACAT 2100
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 684 yrHisGlnGlyArgValIleAspCysSerGlyAlaHisValValLeuAsp 700

2101 GATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTC 2150
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 701 AspAspThrAspValGlyTyrValGluAspGlyThrProCysGlyProSe 717

2151 TATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATACGA 2200
     |||||||||||||||||||||||||||||||||||||||||||||||| ||
 717 rMetMetCysLeuAspArgLysCysLeuGlnIleGlnAlaLeuAsnMetS 734

2201 GCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTG 2250
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 734 erSerCysProLeuAspSerLysGlyLysValCysSerGlyHisGlyVal 750

 251 TGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGA 2300
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 751 CysSerAsnGluAlaThrCysIleCysAspPheThrTrpAlaGlyThrAs 767

2301 TTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAAGGATGAAG 2350
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 767 pCysSerIleArgAspProValArgAsnLeuHisProProLysAspGluG 784

2351 GACCCAAGGGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGGT 2400
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 784 lyProLysGlyProSerAlaThrAsnLeuIleIleGlySerIleAlaGly 800

2401 GCCATCCTGGTAGCAGCTATTGTCCTTGGGGGCACAGGCTGGGGATTTAA 2450
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 801 AlaIleLeuValAlaAlaIleValLeuGlyGlyThrGlyTrpGlyPheLy 817

2451 AAATGTCAAGAAGAGAAGGTTCGATCCTACTCAGCAAGGCCCCATC 2496
     |||||||||||||||||||||||||||||||||||||||||||||
 817 sAsnValLysLysArgArgPheAspProThrGlnGlnGlyProIle 832
```

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY54457

```
seq_documentation_block:
ID   AAY54457 standard; Protein; 696 AA.
XX
AC   AAY54457;
XX
DT   25-APR-2000  (first entry)
XX
DE   Amino acid sequence of a disintegrin homologue designated zdintl.
XX
KW   Human; disintegrin homologue; zdintl; cardiac myocyte; adipocyte;
KW   gene therapy; cell-cell interaction; chondro sacroma; atherosclerosis;
KW   Alzheimer's disease; restenosis; ischemic reperfusion; obesity;
KW   intimal hyperplasia; tumour; platelet aggregation; apoptosis;
KW   neurogenesis; myogenesis; arthritis; myogenesis; neurogenesis;
KW   connective tissue disorder; chondrogenesis; tumour proliferation;
KW   inflammation.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..163
FT                   /note= "propeptide sequence"
FT   Domain          164..382
FT                   /note= "protease sequence; this sequence is
FT                        specifically claimed in claim 4"
FT   Domain          383..464
FT                   /note= "disintegrin sequence; this sequence is
FT                        specifically claimed in claim 5"
FT   Misc-difference 437..450
FT                   /note= "this sequence is specifically claimed in claim 2"
FT   Region          438..449
FT                   /note= "disintegrin loop"
FT   Region          443..445
FT                   /note= "RGD binding loop"
FT   Domain          465..696
FT                   /note= "cysteine-rich domain; this sequence is
FT                        specifically claimed in claim 6"
XX
PN   WO200002912-A2
XX
PD   20-JAN-2000.
XX
PF   09-JUL-1999;   99WO-US15638.
XX
PR   10-JUL-1998;   98US-0113883.
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Sheppard PO,  Baindur N,  Deisher TA,  Bishop PD;
XX
DR   WPI; 2000-160898/14.
DR   N-PSDB; AAZ45826.
XX
PT   Polypeptide useful in modulating cell-cell interaction in tissues of
PT   heart, brain, spinal cord and treating chondro sarcoma,
PT   atherosclerosis, restenosis, obesity, intimal hyperplasia and tumors -
XX
PS   Claim 7; Page 127-129; 132pp; English.
XX
CC   The present sequence represents a human disintegrin homologue, designated
CC   zdintl. The zdintl polypeptide is a cardiac myocyte proliferation and
CC   differentiation stimulator, as well as an adipocyte proliferation and
CC   differentiation inhibitor. Polynucleotides encoding zdintl are used in
CC   gene therapy. The zdintl polypeptide is useful in modulating cell-cell
CC   interactions of cells derived from tissues of heart, brain, spinal cord
CC   and skeletal muscle. It is useful in treating and diagnosing chondro
CC   sacroma, atherosclerosis, Alzheimer's disease, restenosis, ischemic
CC   reperfusion, obesity, intimal hyperplasia and tumors of heart, brain,
CC   and spinal cord. The zdintl polypeptide is also useful in identifying
CC   its new family members, antagonists, agonists and antibodies.
CC   Antagonists, antibodies and fusion proteins of zdintl are useful in
CC   inhibiting platelet aggregation, apoptosis, neurogenesis and myogenesis.
CC   Agonists and antagonists are useful in studying cell-cell interactions,
CC   arthritis, myogenesis, neurogenesis, connective tissue disorders,
CC   chondrogenesis, tumour proliferation and suppression, extracellular
CC   matrix proteins, repair and remodelling of ischemic reperfusion,
CC   inflammation, and apoptosis.
XX
```

```
SQ   Sequence   696 AA;

alignment_scores:
            Quality: 3621.00              Length:    656
              Ratio:  5.537                Gaps:      0
 Percent Similarity:  99.695   Percent Identity: 99.695

alignment_block:
 US-09-634-252A-2 x AAY54457   ..

 Align seg 1/1  to: AAY54457  from: 1  to: 696

     394 CTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCT 443
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       9 LeuAspThrLysAlaArgHisGlnGlnLysHisAsnLysAlaValHisLe 25

     444 GGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTG 493
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      25 uAlaGlnAlaSerPheGlnIleGluAlaPheGlySerLysPheIleLeuA 42

     494 ACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCAC 543
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      42 spLeuIleLeuAsnAsnGlyLeuLeuSerSerAspTyrValGluIleHis 58

     544 TACGAAAATGGGAAACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTA 593
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      59 TyrGluAsnGlyLysProGlnTyrSerLysGlyGlyGluHisCysTyrTy 75

     594 CCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCT 643
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      75 rHisGlySerIleArgGlyValLysAspSerLysValAlaLeuSerThrC 92

     644 GCAATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATA 693
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      92 ysAsnGlyLeuHisGlyMetPheGluAspAspThrPheValTyrMetIle 108

     694 GAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCACATAT 743
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     109 GluProLeuGluLeuValHisAspGluLysSerThrGlyArgProHisIl 125

     744 AATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCA 793
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     125 eIleGlnLysThrLeuAlaGlyGlnTyrSerLysGlnMetLysAsnLeuT 142

     794 CTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTG 843
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     142 hrMetGluArgGlyAspGlnTrpProPheLeuSerGluLeuGlnTrpLeu 158

     844 AAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAAT 893
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     159 LysArgArgLysArgAlaValAsnProSerArgGlyIlePheGluGluMe 175

     894 GAAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGC 943
         ||||||||||||||||||||    ||||||||||||||||||||||||||
     175 tLysTyrLeuGluLeuMetIleGlyAsnAspHisLysThrTyrLysLysH 192

     944 ATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAAC 993
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     192 isArgSerSerHisAlaHisThrAsnAsnPheAlaLysSerValValAsn 208

     994 CTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGT 1043
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     209 LeuValAspSerIleTyrLysGluGlnLeuAsnThrArgValValLeuVa 225

    1044 GGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACC 1093
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     225 lAlaValGluThrTrpThrGluLysAspGlnIleAspIleThrThrAsnP 242

    1094 CTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATTAAGCAG 1143
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     242 roValGlnMetLeuHisGluPheSerLysTyrArgGlnArgIleLysGln 258

    1144 CATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAG 1193
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     259 HisAlaAspAlaValHisLeuIleSerArgValThrPheHisTyrLysAr 275

    1194 AAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTG 1243
         ||||||||||||||||||  ||||||||||||||||||||||||||||
     275 gSerSerLeuSerTyrPheGluGlyValCysSerArgThrArgGlyValG 292

    1244 GTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAG 1293
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     292 lyValAsnGluTyrGlyLeuProMetAlaValAlaGlnValLeuSerGln 308

    1294 AGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCC 1343
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     309 SerLeuAlaGlnAsnLeuGlyIleGlnTrpGluProSerSerArgLysPr 325

    1344 AAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAG 1393
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     325 oLysCysAspCysThrGluSerTrpGlyGlyCysIleMetGluGluThrG 342

    1394 GGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGA 1443
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     342 lyValSerHisSerArgLysPheSerLysCysSerIleLeuGluTyrArg 358

    1444 GACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAA 1493
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     359 AspPheLeuGlnArgGlyGlyGlyAlaCysLeuPheAsnArgProThrLy 375

    1494 GCTATTTGAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGG 1543
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     375 sLeuPheGluProThrGluCysGlyAsnGlyTyrValGluAlaGlyGluG 392

    1544 AGTGTGATTGTGGTTTTCATGTGGAATGCTATGGATTATGCTGTAAGAAA 1593
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     392 luCysAspCysGlyPheHisValGluCysTyrGlyLeuCysCysLysLys 408

    1594 TGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAA 1643
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     409 CysSerLeuSerAsnGlyAlaHisCysSerAspGlyProCysCysAsnAs 425

    1644 TACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGA 1693
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     425 nThrSerCysLeuPheGlnProArgGlyTyrGluCysArgAspAlaValA 442

    1694 ACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCA 1743
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     442 snGluCysAspIleThrGluTyrCysThrGlyAspSerGlyGlnCysPro 458

    1744 CCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCG 1793
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     459 ProAsnLeuHisLysGlnAspGlyTyrAlaCysAsnGlnAsnGlnGlyAr 475

    1794 CTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCT 1843
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     475 gCysTyrAsnGlyGluCysLysThrArgAspAsnGlnCysGlnTyrIleT 492

    1844 GGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAAT 1893
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     492 rpGlyThrLysAlaAlaGlySerAspLysPheCysTyrGluLysLeuAsn 508

    1894 ACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGAT 1943
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     509 ThrGluGlyThrGluLysGlyAsnCysGlyLysAspGlyAspArgTrpIl 525

    1944 TCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATC 1993
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     525 eGlnCysSerLysHisAspValPheCysGlyPheLeuLeuCysThrAsnL 542

    1994 TTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACT 2043
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     542 euThrArgAlaProArgIleGlyGlnLeuGlnGlyGluIleIleProThr 558

    2044 TCCTTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGT 2093
```

```
     559 SerPheTyrHisGlnGlyArgValIleAspCysSerGlyAlaHisValVa 575

    2094 TTTAGATGATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTG 2143
     575 lLeuAspAspAspThrAspValGlyTyrValGluAspGlyThrProCysG 592

    2144 GCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTA 2193
     592 lyProSerMetMetCysLeuAspArgLysCysLeuGlnIleGlnAlaLeu 608

    2194 AATATGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCA 2243
     609 AsnMetSerSerCysProLeuAspSerLysGlyLysValCysSerGlyHi 625

     244 TGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAG 2293
     625 sGlyValCysSerAsnGluAlaThrCysIleCysAspPheThrTrpAlaG 642

    2294 GGACAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAAG 2343
     642 lyThrAspCysSerIleArgAspProValArgAsnLeuHisProProLys 658

    2344 GATGAAGGACCCAAGGGT 2361
     659 AspGluGlyProLysGly 664
```

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY30208

seq_documentation_block:

```
ID   AAY30208 standard; Protein; 867 AA.
XX
AC   AAY30208;
XX
DT   29-OCT-1999  (first entry)
XX
DE   Amino acid sequence of the human SVPH3-13 protein.
XX
KW   SVPH3-13; metalloproteinase-disintegrin; human chromosome 2;
KW   human chromosome 7; malignant hyperthermia susceptibility;
KW   Zellweger syndrome; neonatal adrenoleukodystrophy;
KW   infantile Refsum disease; progressive familial intrahepatic colchestatis;
KW   mucopolysaccharidosis VII; split hand/foot malformation;
KW   arrhythymogenic right ventricular displasia-4; Coppock-like cataract;
KW   insulin dependent diabetes mellitus-12; lamellar type ichthyosis;
     transient neonatal myasthenia gravis; congenital aculeiform cataract;
     juvenile amyotrophic lateral sclerosis;
     familial paroxysmal choreoathetosis;
     Finnish lethal neonatal metabolic syndrome.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 743
FT                   /note= "encoded by GAC"
XX
PN   WO9941388-A2.
XX
PD   19-AUG-1999.
XX
PF   11-FEB-1999;   99WO-US03016.
XX
PR   11-FEB-1998;   98US-0074310.
XX
PA   (IMMV ) IMMUNEX CORP.
XX
PI   Cerretti DP;
XX
DR   WPI; 1999-527371/44.
DR   N-PSDB; AAZ10208.
XX
PT   DNA encoding the SVPH3-13 and SVPH3-17 proteins for detecting
PT   disease corresponding to chromosome 7, e.g. Zellweger syndrome
XX
PS   Claim 13; Page 10-11; 82pp; English.
XX
CC   The present sequence represents a protein designated SVPH3-13, which
CC   is a member of the metalloproteinase-disintegrin family. The
CC   specification also describes SVPH3-17 proteins. Both sequences can
CC   be used to identify human chromosome 2 or 7, and to map genes on these
CC   two chromosomes, and also to identify genes associated with certain
CC   diseases, syndromes, or other human conditions associated with human
CC   chromosome 2 or 7. The disease that correspond to chromosome 7 include
CC   malignant hyperthermia susceptibility, Zellweger syndrome, neonatal
CC   adrenoleukodystrophy, infantile Refsum disease, progressive familial
CC   intrahepatic colchestatis, mucopolysaccharidosis VII, and split
CC   hand/foot malformation. Diseases associated with chromosome 2 include
CC   arrhythymogenic right ventricular displasia-4, insulin dependent diabetes
CC   mellitus-12, transient neonatal myasthenia gravis, juvenile amyotrophic
CC   lateral sclerosis, congenital aculeiform cataract, Coppock-like cataract,
CC   lamellar type ichthyosis, familial paroxysmal choreoathetosis, and
CC   Finnish lethal neonatal metabolic syndrome.
XX
SQ   Sequence   867 AA;
```

alignment_scores:

```
           Quality: 1985.00              Length:    746
             Ratio:  3.564                Gaps:     12
 Percent Similarity:  74.665   Percent Identity: 50.268
```

alignment_block:

```
 US-09-634-252A-2 x AAY30208   ..

 Align seg 1/1  to: AAY30208  from: 1  to: 867

     319 ATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACATCAACCA 368
       6 ValGluArgGlnSerIleValProLeuArgLeuIleTyrArgSerGlyGl 22

     369 AGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGC 418
      22 yGluAspGluSerArgHisAspAlaLeuAspThrArgValArgGlyAspL 39

     419 AAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAA 468
      39 euGlyGlyArgGlnLeuThrHisValAspGlnAlaSerPheGlnValAsp 55

     469 GCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGTT 518
      56 AlaPheGlyThrSerPheIleLeuAspValValLeuAsnHisAspLeuLe 72

     519 GTCTTCTGATTATGTGGAGATTCACTACGAAAAT...GGGAAACCACAGT 565
      72 uSerSerGluTyrIleGluArgHisIleGluHisGlyGlyLysThrValG 89

     566 ACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGCGTC 615
      89 luValLysGlyGlyGluHisCysTyrTyrGlnGlyHisIleArgGlyAsn 105

     616 AAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTT 665
     106 ProAspSerPheValAlaLeuSerThrCysHisGlyLeuHisGlyMetPh 122

     666 TGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATG 715
     122 eTyrAspGlyAsnHisThrTyrLeuIleGluPro.............G 134

     716 ATGAGAAAAGCACAGGTCGACCA...........CATATAATCCAGAAA 753
     134 luGluAsnAspThrThrGlnGluAspPheHisPheHisSerValTyrLys 150

     754 ACCTTGGCAGGACAGTATTCTAAG.................CAAAT 782
     151 SerArgLeuPheGluPheSerLeuAspAspLeuProSerGluPheGlnGl 167
```

```
 783 GAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAAT 832
 167 nValAsnIleThrProSerLys..............PheIleLeuLysP 179

 833 TACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGTATA 882
 179 roArgProLysArgSerLysArgGlnLeuArgArgTyrProArgAsnVal 195

 883 TTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAAC 932
 196 GluGluGluThrLysTyrIleGluLeuMetIleValAsnAspHisLeuMe 212

 933 GTATAAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGT 982
 212 tPheLysLysHisArgLeuSerValValHisThrAsnThrTyrAlaLysS 229

 983 CCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGG 1032
 229 erValValAsnMetAlaAspLeuIleTyrLysAspGlnLeuLysThrArg 245

1033 GTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACAT 1082
1246 IleValLeuValAlaMetGluThrTrpAlaThrAspAsnLysPheAlaIl 262

1083 CACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGC 1132
 262 eSerGluAsnProLeuIleThrLeuArgGluPheMetLysTyrArgArgA 279

1133 GC...ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACA 1179
 279 spPheIleLysGluLysSerAspAlaValHisLeuPheSerGlySerGln 295

1180 TTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCG 1229
 296 PheGluSerSerArgSerGlyAlaAlaTyrIleGlyGlyIleCysSerLe 312

1230 CACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCAC 1279
 312 uLeuLysGlyGlyGlyValAsnGluPheGlyLysThrAspLeuMetAlaV 329

1280 AAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCT 1329
 329 alThrLeuAlaGlnSerLeuAlaHisAsnIleGlyIle..... 341

1330 TCTAGCAGAAAGCCAAAA..............TGTGACTGCACAGAATC 1364
 342 IleSerAspLysArgLysLeuAlaSerGlyGluCysLysCysGluAspTh 358

1365 CTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAAT 1414
 358 rTrpSerGlyCysIleMetGlyAspThrGlyTyrTyrLeuProLysLysP 375

1415 TTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGT 1464
1375 heThrGlnCysAsnIleGluGluTyrHisAspPheLeuAsnSerGlyGly 391

1465 GGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATG 1514
 392 GlyAlaCysLeuPheAsnLysProSerLysLeuLeuAspProProGluCy 408

1515 TGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATG 1564
 408 sGlyAsnGlyPheIleGluThrGlyGluGluCysAspCysGlyThrProA 425

1565 TGGAATGCTAT.........GGATTATGCTGTAAGAAATGTTCCCTCTCC 1605
 425 laGluCysValLeuGluGlyAlaGluCysCysLysLysCysThrLeuThr 441

1606 AACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCT 1655
 442 GlnAspSerGlnCysSerAspGlyLeuCysCysLys......LysCysLy 456

1656 TTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATA 1705
```

```
 456 sPheGlnProMetGlyThrValCysArgGluAlaValAsnAspCysAspI 473

1706 TTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCAT 1755
 473 leArgGluThrCysSerGlyAsnSerSerGlnCysAlaProAsnIleHis 489

1756 AAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGG 1805
 490 LysMetAspGlyTyrSerCysAspGlyValGlnGlyIleCysPheGlyGl 506

1806 CGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGG 1855
 506 yArgCysLysThrArgAspArgGlnCysLysTyrIleTrpGlyGlnLysV 523

1856 CTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACT 1905
 523 alThrAlaSerAspLysTyrCysTyrGluLysLeuAsnIleGluGlyThr 539

1906 GAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAA 1955
 540 GluLysGlyAsnCysGlyLysAspLysAspThrTrpIleGlnCysAsnLy 556

1956 ACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTC 2005
 556 sArgAspValLeuCysGlyTyrLeuLeuCysThrAsnIleGlyAsnIleP 573

2006 CACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCAT 2055
 573 roArgLeuGlyGluLeuAspGlyGluIleThrSerThrLeuValValGln 589

2056 CAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGA 2105
 590 GlnGlyArgThrLeuAsnCysSerGlyGlyHisValLysLeuGluGluAs 606

2106 TACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGA 2155
 606 pValAspLeuGlyTyrValGluAspGlyThrProCysGlyProGlnMetM 623

2156 TGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGC 2205
 623 etCysLeuGluHisArgCysLeuProValAlaSerPheAsnPheSerThr 639

2206 TGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAG 2255
 640 CysLeuSerSerLysGluGlyThrIleCysSerGlyAsnGlyValCysSe 656

2256 TAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCA 2305
 656 rAsnGluLeuLysCysValCysAsnArgHisTrpIleGlySerAspCys. 672

2306 GTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAAGGATGAAGGACCC 2355
 673 ..................AsnThrTyrPheProHisAsnAspAspAla 682

2356 AAG.....................GGTCCTAGTGCCACCAATCTCAT 2381
 683 LysThrGlyIleThrLeuSerGlyAsnGlyValAlaGlyThrAsnIleIl 699

2382 AATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCTATTGTCCTTGGGG 2431
 699 eIleGlyIleIleAlaGlyThrIleLeuValLeuAlaLeuIleLeuGlyI 716

2432 GCACAGGCTGGGGATTTAAAAATGTCAAGAAGAGAAGG 2469
 716 leThrAlaTrpGlyTyrLysAsnTyrArgGluGlnArg 728

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY25119

seq_documentation_block:
ID   AAY25119 standard; Protein; 823 AA.
XX
AC   AAY25119;
```

```
XX
DT   26-AUG-1999  (first entry)
XX
DE   Human MDC2-beta protein.
XX
KW   Metalloproteinase-like-disintegrin-like cysteine rich protein; human;
KW   MDC2-alpha; MDC2-beta; MDC3; medical; treatment; diagnosis.
XX
OS   Homo sapiens.
XX
PN   JP11155574-A.
XX
PD   15-JUN-1999.
XX
PF   01-DEC-1997;   97JP-0330020.
XX
PR   01-DEC-1997;   97JP-0330020.
XX
PA   (EISA ) EISAI CO LTD.
XX
DR   WPI; 1999-398071/34.
DR   N-PSDB; AAX78437.
XX
PT   New protein belonging to MDC gene family  - useful in medical
PT   treatment and diagnosis
XX
PS   Disclosure; Page 10-13; 17pp; Japanese.
XX
CC   This invention describes novel human MDC2-alpha, MDC2-beta and MDC3
CC   proteins and their encoding nucleic acids. The products of the invention
CC   are useful in medical treatment and diagnosis.
XX
SQ   Sequence   823 AA;


alignment_scores:
            Quality: 1983.50              Length:    817
              Ratio:  3.432                Gaps:     14
 Percent Similarity:  70.747   Percent Identity: 47.491

alignment_block:
 US-09-634-252A-2 x AAY25119   ..

 Align seg 1/1  to: AAY25119  from: 1  to: 823

     106 AGCGCCCCGGCCCGCACGCCGCCCTGCCGCCTGCTTCTCGTCCTTCTCCT 155
         :::|||   |||   :::   |||      ||||||   ||||||
       3 AlaAlaValAlaValSerValProPheLeuLeuLeuCysValLeuGlyTh 19

     156 GCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCGCGCCTGGGGGGCTGCTG 205
             ||||||      :::                      |||   |||:
      19 rCysProProAlaArgCys.....................GlyGlnAlaG 29

     206 CGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAAAAAAATTTGGGAGTC 255
         ::   ::::::            |||      :::::::|||
      29 lyAspAlaSerLeuMetGluLeuGluLysArgLysGluAsn......... 42

     256 CTGGCAGATGAAGACAATACATTGCAACAGAATAGCAGCAGTAATATCAG 305

      42 ................................................. 42

     306 TTACAGCAATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATAT 355
                ::   ::::::::::::      :::|||   ||||||||||
      43 .......ArgPheValGluArgGlnSerIleValProLeuArgLeuIleT 57

     356 ATTACATCAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAG 405
         ||      :::   ::::::||||||   :::   :::||||||||:::
      57 yrArgSerGlyGlyGluAspGluSerArgHisAspAlaLeuAspThrArg 73

     406 GCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAG 455
         :::|||   :::            :::   :::|||:::   ||||||||
      74 ValArgGlyAspLeuGlyGlyProGlnLeuThrHisValAspGlnAlaSe 90

     456 CTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGA 505
         |||||||::::::|||||||||::::::|||||||||||:::::::||||
      90 rPheGlnValAspAlaPheGlyThrSerPheIleLeuAspValValLeuA 107

     506 ACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAAAT... 552
         ||:::   |||||||||||||:::|||:::|||   |||   |||:::
     107 snHisAspLeuLeuSerSerGluTyrIleGluArgHisIleGluHisGly 123

     553 GGGAAACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAG 602
         ||||||            |||||||||||||||||||||||:::|||
     124 GlyLysThrValGluValLysGlyGlyGluHisCysTyrTyrGlnGlyHi 140

     603 CATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGAC 652
          |||||||||      ||||||   ||||||||||||||||||:::||||
     140 sIleArgGlyAsnProAspSerPheValAlaLeuSerThrCysHisGlyL 157

     653 TTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTA 702
         ||||||||||||||   |||   :::   :::|||:::|||||||||
     157 euHisGlyMetPheTyrAspGlyAsnHisThrTyrLeuIleGluPro... 172

     703 GAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCA.............CA 740
                     :::|||::::::|||   :::                ||
     173 .............GluGluAsnAspThrThrGlnGluAspPheHisPheHi-185

     741 TATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAG.............. 777
         |   :::   |||:::          ::::::|||
     185 sSerValTyrLysSerArgLeuPheGluPheSerLeuAspAspLeuProS 202

     778 ........CAAATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCC 819
                 |||:::   |||:::|||   ::::::
     202 erGluPheGlnGlnValAsnIleThrProSerLys.............. 213

     820 TTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCC 869
         |||:::   :::   :::       :::   :::::::::      :::
     214 PheIleLeuLysProArgProLysArgSerLysArgGlnLeuArgArgTy 230

     870 ATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTA 919
             |||::::::   ||||||   ||||||:::||||||||||||||||
     230 rProArgAsnValGluGluGluThrLysTyrIleGluLeuMetIleValA 247

     920 ATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAAC 969
         ||||||||      :::||||||||||||||   |||   :::||||||||||
     247 snAspHisLeuMetPheLysLysHisArgLeuSerValValHisThrAsn 263

     970 AACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCA 1019
         ::::::|||||||||||||||||:::::::|||   ||||||||||:::||
     264 ThrTyrAlaLysSerValValAsnMetAlaAspLeuIleTyrLysAspGl 280

    1020 GCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGG 1069
         ||||:::||||||:::||||||||||||:::||||||||||:::      :
     280 nLeuLysThrArgIleValLeuValAlaMetGluThrTrpAlaThrAspA 297

    1070 ATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCA 1119
         ::::::::   |||:::   ||||||::       |||:::||||||
     297 snLysPheAlaIleSerGluAsnProLeuIleThrLeuArgGluPheMet 313

    1120 AAATACCGGCAGCGC...ATTAAGCAGCATGCTGATGCTGTGCACCTCAT 1166
         |||||||||:::      ||||||:::   :::|||||||||||||::
     314 LysTyrArgArgAspPheIleLysGluLysSerAspAlaValHisLeuPh 330

    1167 CTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAG 1216
         :|||          |||:::   :::||||||:::   :::|||:::||||
     330 eSerGlySerGlnPheGluSerSerArgSerGlyAlaAlaTyrIleGlyG 347

    1217 GTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCA 1266
         ||:::||||||      :::|||   |||||||||||:::|||
     347 lyIleCysSerLeuLeuLysGlyGlyGlyValAsnGluPheGlyLysThr 363

    1267 ATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAAT 1316
               :::|||   :::|||:::||||||||||||:::|||:::||||
     364 AspLeuMetAlaValThrLeuAlaGlnSerLeuAlaHisAsnIleGlyIl 380

    1317 CCAATGGGAACCTTCTAGCAGAAAGCCAAAA.............TGTG 1351
```

```
 380 e............IleSerAspLysArgLysLeuAlaSerGlyGluCysL 393

1352 ACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCC 1401
 393 ysCysGluAspThrTrpSerGlyCysIleMetGlyAspThrGlyTyrTyr 409

1402 CATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTT 1451
 410 LeuProLysLysPheThrGlnCysAsnIleGluGluTyrHisAspPheLe 426

1452 ACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTG 1501
 426 uAsnSerGlyGlyGlyAlaCysLeuPheAsnLysProSerLysLeuLeuA 443

1502 AGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGAT 1551
 443 spProProGluCysGlyAsnGlyPheIleGluThrGlyGluGluCysAsp 459

1552 TGTGGTTTTCATGTGGAATGCTAT.........GGATTATGCTGTAAGAA 1592
  60 CysGlyThrProAlaGluCysValLeuGluGlyAlaGluCysCysLysLy 476

 93 ATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACA 1642
 476 sCysThrLeuThrGlnAspSerGlnCysSerAspGlyLeuCysCysLys. 492

1643 ATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTG 1692
 493 .....LysCysLysPheGlnProMetGlyThrValCysArgGluAlaVal 507

1693 AACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCC 1742
 508 AsnAspCysAspIleArgGluThrCysSerGlyAsnSerSerGlnCysAl 524

1743 ACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCC 1792
 524 aProAsnIleHisLysMetAspGlyTyrSerCysAspGlyValGlnGlyI 541

1793 GCTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATC 1842
 541 leCysPheGlyGlyArgCysLysThrArgAspArgGlnCysLysTyrIle 557

1843 TGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAA 1892
 558 TrpGlyGlnLysValThrAlaSerAspLysTyrCysTyrGluLysLeuAs 574

1893 TACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGA 1942
 574 nIleGluGlyThrGluLysGlyAsnCysGlyLysAspLysAspThrTrpI 591

1943 TTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAAT 1992
 591 leGlnCysAsnLysArgAspValLeuCysGlyTyrLeuLeuCysThrAsn 607

1993 CTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAAC 2042
 608 IleGlyAsnIleProArgLeuGlyGluLeuAspGlyGluIleThrSerTh 624

2043 TTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAG 2092
 624 rLeuValValGlnGlnGlyArgThrLeuAsnCysSerGlyGlyHisValL 641

2093 TTTTAGATGATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGT 2142
 641 ysLeuGluGluAspValAspLeuGlyTyrValGluAspGlyThrProCys 657

2143 GGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCT 2192
 658 GlyProGlnMetMetCysLeuGluHisArgCysLeuProValAlaSerPh 674

2193 AAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCC 2242
```

```
 674 eAsnPheSerThrCysLeuSerSerLysGluGlyThrIleCysSerGlyA 691

2243 ATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCA 2292
 691 snGlyValCysSerAsnGluLeuLysCysValCysAsnArgHisTrpIle 707

2293 GGGACAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAA 2342
 708 GlySerAspCys.....AsnThrTyrPheProHi 717

2343 GGATGAAGGACCCAAG.....GGTCCTAGTG 2368
 717 sAsnAspAspAlaLysThrGlyIleThrLeuSerGlyAsnGlyValAlaG 734

2369 CCACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCT 2418
 734 lyThrAsnIleIleIleGlyIleIleAlaGlyThrIleLeuValLeuAla 750

2419 ATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAAATGTCAAGAAGAGAAG 2468
 751 LeuIleLeuGlyIleThrAlaTrpGlyTyrLysAsnTyrArgGluGlnAr 767

2469 G 2469
 767 g 767
```

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY25118

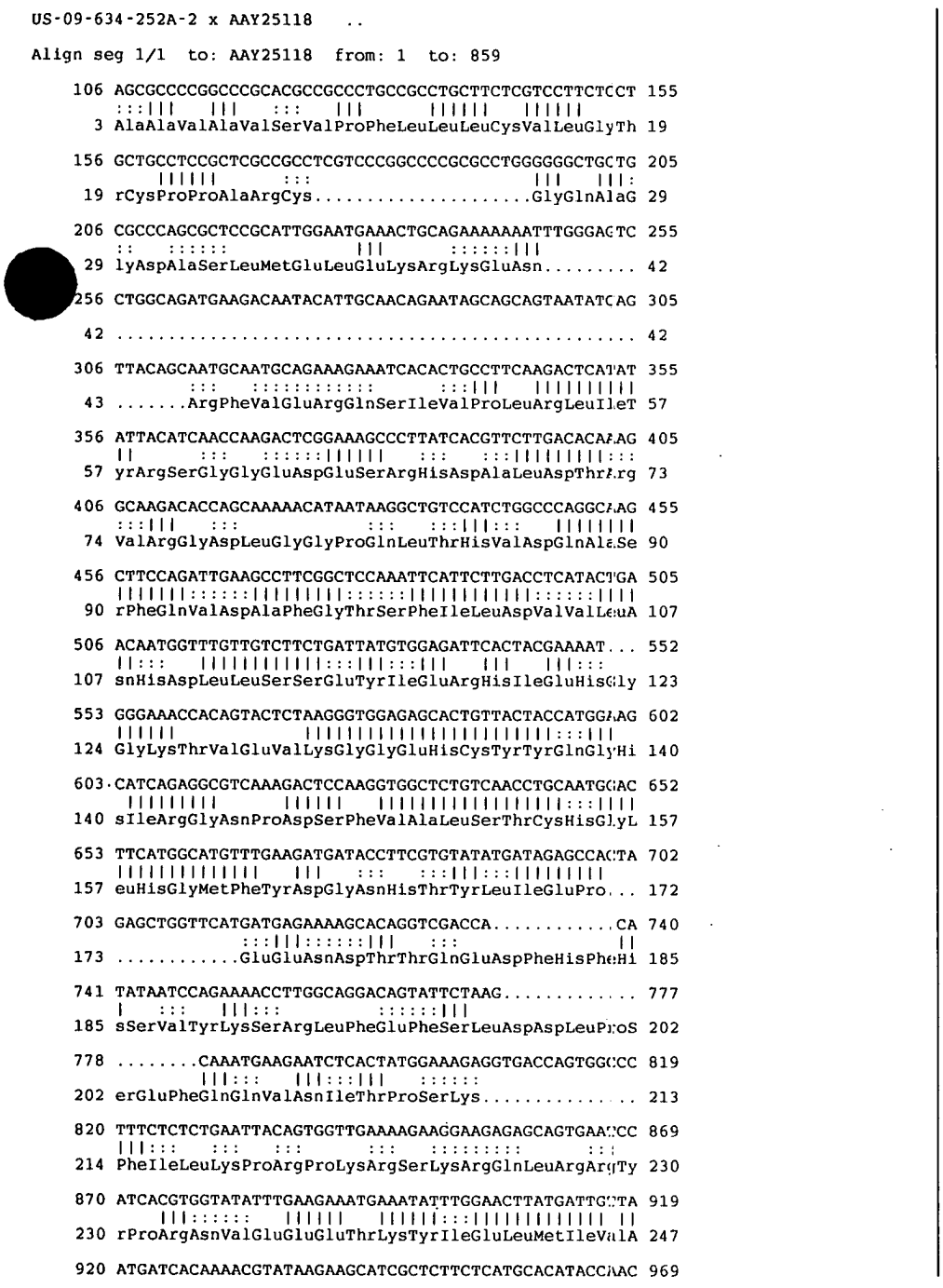
seq_documentation_block:
ID AAY25118 standard; Protein; 859 AA.
XX
AC AAY25118;
XX
DT 26-AUG-1999 (first entry)
XX
DE Human MDC2-alpha protein.
XX
KW Metalloproteinase-like-disintegrin-like cysteine rich protein; human;
KW MDC2-alpha; MDC2-beta; MDC3; medical; treatment; diagnosis.
XX
OS Homo sapiens.
XX
PN JP11155574-A.
XX
PD 15-JUN-1999.
XX
PF 01-DEC-1997; 97JP-0330020.
XX
PR 01-DEC-1997; 97JP-0330020.
XX
PA (EISA) EISAI CO LTD.
XX
DR WPI; 1999-398071/34.
DR N-PSDB; AAX78436.
XX
PT New protein belonging to MDC gene family - useful in medical
PT treatment and diagnosis
XX
PS Disclosure; Page 6-10; 17pp; Japanese.
XX
CC This invention describes novel human MDC2-alpha, MDC2-beta and MDC3
CC proteins and their encoding nucleic acids. The products of the invention
CC are useful in medical treatment and diagnosis.
XX
SQ Sequence 859 AA;

alignment_scores:
Quality: 1983.50 Length: 817
Ratio: 3.432 Gaps: 14
Percent Similarity: 70.747 Percent Identity: 47.491

alignment_block:

US-09-634-252A-2 x AAY25118 ..

Align seg 1/1 to: AAY25118 from: 1 to: 859



```
 541 leCysPheGlyGlyArgCysLysThrArgAspArgGlnCysLysTyrIle 557

1843 TGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAA 1892
     ||||||   |||:::::::::|||||||||:::||||||||||||||||
 558 TrpGlyGlnLysValThrAlaSerAspLysTyrCysTyrGluLysLeuAs 574

1893 TACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGA 1942
     |   ||||||||||||||||||||||||||||||   |||   ||||
 574 nIleGluGlyThrGluLysGlyAsnCysGlyLysAspLysAspThrTrpI 591

1943 TTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAAT 1992
     ||||||||:::|||:::||||||:::||||||:::|||||||||||||
 591 leGlnCysAsnLysArgAspValLeuCysGlyTyrLeuLeuCysThrAsn 607

1993 CTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAAC 2042
     :::   :::   ||||||:::|||:::|||:::|||||||||      ||
 608 IleGlyAsnIleProArgLeuGlyGluLeuAspGlyGluIleThrSerTh 624

2043 TTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAG 2092
     |         :::||||||||||:::::::::||||||||||:::|||||||
 624 rLeuValValGlnGlnGlyArgThrLeuAsnCysSerGlyGlyHisValL 641

2093 TTTTAGATGATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGT 2142
       |||:::::::|||:::|||:::||||||||||||||||||||||||||
 641 ysLeuGluGluAspValAspLeuGlyTyrValGluAspGlyThrProCys 657

2143 GGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCT 2192
     ||||||:::||||||||||||:::::::::||||||   :::   :::::
 658 GlyProGlnMetMetCysLeuGluHisArgCysLeuProValAlaSerPh 674

2193 AAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCC 2242
     :|||:::|||:::|||     :::::::::|||   :::|||||||||:
 674 eAsnPheSerThrCysLeuSerSerLysGluGlyThrIleCysSerGlyA 691

2243 ATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCA 2292
     ::|||||||||||||||||      |||:::|||:::      |||
 691 snGlyValCysSerAsnGluLeuLysCysValCysAsnArgHisTrpIle 707

2293 GGGACAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAA 2342
     |||:::||||||                  |||   :::   |||
 708 GlySerAspCys.....................AsnThrTyrPheProHi 717

2343 GGATGAAGGACCCAAG.....................GGTCCTAGTG 2368
      ::::::      |||                     |||   ::::
 717 sAsnAspAspAlaLysThrGlyIleThrLeuSerGlyAsnGlyValAlaG 734

2369 CCACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCT 2418
     ::|||||||:::|||||||||   |||||||||:::|||||||||   |||
 734 lyThrAsnIleIleIleGlyIleIleAlaGlyThrIleLeuValLeuAla 750

2419 ATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAAATGTCAAGAAGAGAAG 2468
     ::::::||||||   |||:::||||||:::||||||   ::::::::::||
  51 LeuIleLeuGlyIleThrAlaTrpGlyTyrLysAsnTyrArgGluGlnAr 767

  69 G 2469
     |
 767 g 767

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:AAR75352

seq_documentation_block:
ID   AAR75352 standard; Protein; 769 AA.
XX
AC   AAR75352;
XX
DT   01-AUG-1995  (first entry)
XX
DE   Human fetal brain MDC protein.
XX
KW   MDC protein; breast cancer; mamma carcinoma; ovary cancer;
KW   chromosome-17.
XX
OS   Homo sapiens.
XX
PN   EP633268-A.
XX
PD   11-JAN-1995.
XX
PF   13-MAY-1994;   94EP-0107487.
XX
PR   14-MAY-1993;   93JP-0136602.
PR   22-SEP-1993;   93JP-0257455.
PR   23-FEB-1994;   94JP-0049904.
PR   12-APR-1994;   94JP-0073328.
XX
PA   (CANC-) CANCER INST.
PA   (EISA ) EISAI CO LTD.
XX
PI   Emi M,  Nakamura Y;
XX
DR   WPI; 1995-038478/06.
DR   N-PSDB; AAQ76122.
XX
PT   Novel MDC protein and DNA encoding it - used to develop prods.
PT   for the study, diagnosis and therapy of cancers, partic. breast
PT   and ovarian cancer
XX
PS   Disclosure; Page 94-99; 123pp; English.
XX
CC   A detailed map of human chromosome-17 was constructed to analyze the
CC   chromosome in breast and ovarian cancer tissues.  2 Novel cDNAs
CC   (given in AAQ76120-21) were obtained from fetal cerebellum that
CC   encoded novel MDC proteins (AAR67758-59, respectively).  5'-RACE and
CC   RT-PCR revealed the sequence given in AAQ76122, encoding the MDC given
CC   in AAR75352.  Sequences common for these clones are given in AAQ76119
CC   and AAR66757, respectively.  A genomic DNA sequence (AAQ76124) for MDC
CC   was obtained from cosmid clone cCI17-904.
XX
SQ   Sequence   769 AA;


alignment_scores:
            Quality: 1978.00              Length:    817
              Ratio:  3.446                Gaps:     17
 Percent Similarity:  70.257   Percent Identity: 48.592

alignment_block:
 US-09-634-252A-2 x AAR75352   ..

 Align seg 1/1  to: AAR75352  from: 1  to: 769

    136 CTGCTTCTCGTCCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCC 185
        |||||||||   ||||||      |||   |||:::::::::::::
     12 LeuLeuLeuSerLeuLeuProThrProGlyLeuGlyThrGlnGlyProAl 28

    186 CCGCGCC......TGGGGGGCT.............GCTGCGCCCAGCGCTC 217
            |||      ||||||:::             ::::::|||:::||||
     28 aGlyAlaLeuArgTrpGlyGlyLeuProGlnLeuGlyGlyProGlyAlaP 45

    218 CGCATTGGAATGAAACTGCAGAAAAAAATTTGGGAGTCCTGGCAGATGAA 267
        ||
     45 ro.............................................. 45

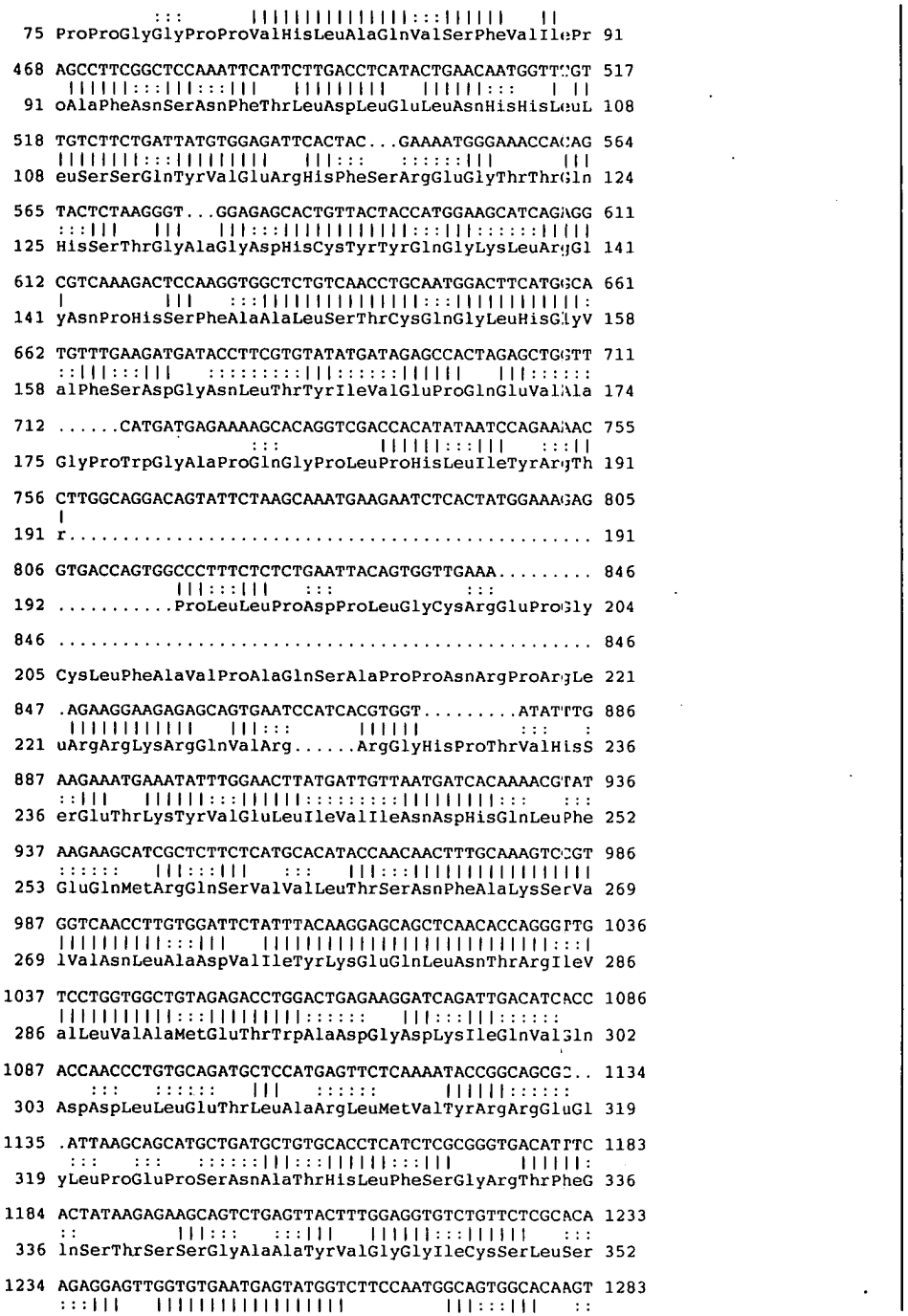
    268 GACAATACATTGCAACAGAATAGCAGCAGTAATATCAGTTACAGCAATGC 317

     45 ..............................................  45

    318 AATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACATCAACC 367
                 |||:::|||   ||||||||||||:::         ::::
     46 ..........GluValThrGluProSerArgLeuValArgGlu...SerS 58

    368 AAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAG 417
        ::   :::|||         :::-  |||||||||::::::|||::::::
     58 erGlyGlyGluValArgLysGlnGlnLeuAspThrArgValArgGlnGlu 74

    418 CAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGA 467
```


75 ProProGlyGlyProProValHisLeuAlaGlnValSerPheValIlePr 91
468 AGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTCGT 517
91 oAlaPheAsnSerAsnPheThrLeuAspLeuGluLeuAsnHisHisLeuL 108
518 TGTCTTCTGATTATGTGGAGATTCACTAC...GAAAATGGGAAACCACAG 564
108 euSerSerGlnTyrValGluArgHisPheSerArgGluGlyThrThrGln 124
565 TACTCTAAGGGT...GGAGAGCACTGTTACTACCATGGAAGCATCAGAGG 611
125 HisSerThrGlyAlaGlyAspHisCysTyrTyrGlnGlyLysLeuArgGl 141
612 CGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCA 661
141 yAsnProHisSerPheAlaAlaLeuSerThrCysGlnGlyLeuHisGlyV 158
662 TGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTT 711
158 alPheSerAspGlyAsnLeuThrTyrIleValGluProGlnGluValAla 174
712 CATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAAC 755
175 GlyProTrpGlyAlaProGlnGlyProLeuProHisLeuIleTyrArgTh 191
756 CTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAG 805
191 r..... 191
806 GTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAA..... 846
192 ProLeuLeuProAspProLeuGlyCysArgGluProGly 204
846 846
205 CysLeuPheAlaValProAlaGlnSerAlaProProAsnArgProArgLe 221
847 .AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT.....ATATTTG 886
221 uArgArgLysArgGlnValArg.....ArgGlyHisProThrValHisS 236
887 AAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTAT 936
236 erGluThrLysTyrValGluLeuIleValIleAsnAspHisGlnLeuPhe 252
937 AAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGT 986
253 GluGlnMetArgGlnSerValValLeuThrSerAsnPheAlaLysSerVa 269
987 GGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTG 1036
269 lValAsnLeuAlaAspValIleTyrLysGluGlnLeuAsnThrArgIleV 286
1037 TCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACC 1086
286 alLeuValAlaMetGluThrTrpAlaAspGlyAspLysIleGlnVal3ln 302
1087 ACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGC.. 1134
303 AspAspLeuLeuGluThrLeuAlaArgLeuMetValTyrArgArgGluGl 319
1135 .ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTC 1183
319 yLeuProGluProSerAsnAlaThrHisLeuPheSerGlyArgThrPheG 336
1184 ACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACA 1233
336 lnSerThrSerSerGlyAlaAlaTyrValGlyGlyIleCysSerLeuSer 352
1234 AGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACAAGT 1283
353 HisGlyGlyGlyValAsnGluTyrGlyAsnMetGlyAlaMetAlaValTh 369
1284 ATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGG...GAACCTT 1330
369 rLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMetMetTrpAsnLysHisA 386
1331 CTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATC 1380
386 rgSerSerAlaGlyAspCysLysCysProAspIleTrpLeuGlyCysIle 402
1381 ATGGAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCAT 1430
403 MetGluAspThrGlyPheTyrLeuProArgLysPheSerArgCysSerIl 419
1431 TTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCA 1480
419 eAspGluTyrAsnGlnPheLeuGlnGluGlyGlySerCysLeuPheA 436
1481 ACAGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATACGTG 1530
436 snLysProLeuLysLeuLeuAspProProGluCysGlyAsnGlyPheVal 452
1531 GAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTAT..... 1575
453 GluAlaGlyGluGluCysAspCysGlySerValGlnGluCysSerArgAl 469
1576 GGATTATGCTGTAAGAAATGTTCCCTCTCCAACGGGGCTCACTGCA 1621
469 aGlyGlyAsnCysCysLysLysCysThrLeuThrHisAspAlaMetCysS 486
1622 GCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGG 1671
486 erAspGlyLeuCysCysArgArg......CysLysTyrGluProArgGly 500
1672 TATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTAC 1721
501 ValSerCysArgGluAlaValAsnGluCysAspIleAlaGluThrCysTh 517
1722 TGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATG 1771
517 rGlyAspSerSerGlnCysProProAsnLeuHisLysLeuAspGlyTyrT 534
1772 CATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGA 1821
534 yrCysAspHisGluGlnGlyArgCysTyrGlyGlyArgCysLysThrArg 550
1822 GACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAA 1871
551 AspArgGlnCysGlnValLeuTrpGlyHisAlaAlaAla......AspAr 565
1872 GTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCG 1921
565 gPheCysTyrGluLysLeuAsnValGluGlyThrGluArgGlySerCysG 582
1922 GGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGT 1971
582 lyArgLysGlySerGlyTrpValGlnCysSerLysGlnAspValLeuCys 598
1972 GGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACT 2021
599 GlyPheLeuLeuCysValAsnIleSerGlyAlaProArgLeuGlyAspLe 615
2022 TCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTG 2071
615 uValGlyAspIleSerSerValThrPheTyrHisGlnGlyLysGluLeuA 632
2072 ACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTAT 2121
632 spCysArgGlyGlyHisValGlnLeuAlaAspGlySerAspLeuSerTyr 648
2122 GTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAA 2171
649 ValGluAspGlyThrAlaCysGlyProAsnMetLeuCysLeuAspHisAr 665

```
    2172 GTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCCA 2221
         :||||||      :::|||:::|||:::|||:::||||||   :::::::
     665 gCysLeuProAlaSerAlaPheAsnPheSerThrCysProGlySerGlyG 682

    2222 AGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGC 2271
         ::   ::::::||||||   |||||||||||||||||||||:::   |||
     682 luArgArgIleCysSerHisHisGlyValCysSerAsnGluGlyLysCys 698

    2272 ATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGT 2321
         ||||||:::      |||:::|||   ||||||||||:::::|||::
     699 IleCysGlnProAspTrpThrGlyLysAspCysSerIleHisAsnProLe 715

    2322 TAGGAACCTTCACCCCCCC...AAGGATGAAGGACCCAAGGGTCCTAGTG 2368
         :   :::      |||      :::   |||      ||||||||||||:
     715 uProThrSerProProThrGlyGluThrGluArgTyrLysGlyProSerG 732

    2369 CCACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCT 2418
         ::||||||:::|||||||||||||||||||||||||:::||||||||||||
     732 lyThrAsnIleIleIleGlySerIleAlaGlyAlaValLeuValAlaAla 748

     419 ATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAAATGTCAAGAAGAGAAG 2468
         ||||||||||||||||||||||||||||||||||||:::::::::   ||
     749 IleValLeuGlyGlyThrGlyTrpGlyPheLysAsnIleArgArgGlyAr 765

    2469 G 2469
         |
     765 g 765

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:AAR67759

seq_documentation_block:
ID   AAR67759 standard; Protein; 670 AA.
XX
AC   AAR67759;
XX
DT   01-AUG-1995  (first entry)
XX
DE   Human fetal brain MDC protein.
XX
KW   MDC protein; breast cancer; mamma carcinoma; ovary cancer;
KW   chromosome-17.
XX
OS   Homo sapiens.
XX
PN   EP633268-A.
XX
PD   11-JAN-1995.
XX
PF   13-MAY-1994;   94EP-0107487.
XX
     14-MAY-1993;   93JP-0136602.
     22-SEP-1993;   93JP-0257455.
     23-FEB-1994;   94JP-0049904.
     12-APR-1994;   94JP-0073328.

PA   (CANC-) CANCER INST.
PA   (EISA ) EISAI CO LTD.
XX
PI   Emi M,  Nakamura Y;
XX
DR   WPI; 1995-038478/06.
DR   N-PSDB; AAQ76121.
XX
PT   Novel MDC protein and DNA encoding it - used to develop prods.
PT   for the study, diagnosis and therapy of cancers, partic. breast
PT   and ovarian cancer
XX
PS   Disclosure; Page 90-94; 123pp; English.
XX
CC   A detailed map of human chromosome-17 was constructed to analyze the
CC   chromosome in breast and ovarian cancer tissues.  2 Novel cDNAs
CC   (given in AAQ76120-21) were obtained from fetal cerebellum that
CC   encoded novel MDC proteins (AAR67758-59, respectively).  5'-RACE and
CC   RT-PCR revealed the sequence given in AAQ76122, encoding the MDC given
CC   in AAR75352.  Sequences common for these clones are given in AAQ76119
CC   and AAR66757, respectively.  A genomic DNA sequence (AAQ76124) for MDC
CC   was obtained from cosmid clone cCI17-904.
XX
SQ   Sequence   670 AA;


alignment_scores:
            Quality: 1862.50            Length:    689
              Ratio:  3.638              Gaps:     13
 Percent Similarity:  74.311   Percent Identity: 51.669

alignment_block:
 US-09-634-252A-2 x AAR67759   ..

 Align seg 1/1  to: AAR67759  from: 1  to: 670

     502 CTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTAC...GA 548
         |||:::::::   |||||||||||:::|||||||||   |||:::   ::
       4 LeuSerHisGlnLeuLeuSerSerGlnTyrValGluArgHisPheSerAr 20

     549 AAATGGGAAACCACAGTACTCTAAGGGT...GGAGAGCACTGTTACTACC 595
         ::::|||      |||:::|||   |||   |||:::|||||||||||||:
      20 gGluGlyThrThrGlnHisSerThrGlyAlaGlyAspHisCysTyrTyrG 37

     596 ATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGC 645
         ::|||::::::||||||         |||   :::|||||||||||||||
      37 lnGlyLysLeuArgGlyAsnProHisSerPheAlaAlaLeuSerThrCys 53

     646 AATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGA 695
         :::|||||||||||||:::|||:::|||   ::::::::::|||:::::||
      54 GlnGlyLeuHisGlyValPheSerAspGlyAsnLeuThrTyrIleValGl 70

     696 GCCACTAGAGCTGGTT......CATGATGAGAAAAGCACAGGTCGACCAC 739
         ||||   |||::::::         :::         ||||
      70 uProGlnGluValAlaGlyProTrpGlyAlaProGlnGlyProLeuProH 87

     740 ATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAAT 789
         ||:::|||   :::|||
      87 isLeuIleTyrArgThr..............................  92

     790 CTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTG 839
                                    |||:::|||   :::
      93 ..........................ProLeuLeuProAspProLeuGl 100

     840 GTTGAAA.....................................  846
             :::
     100 yCysArgGluProGlyCysLeuPheAlaValProAlaGlnSerAlaProP 117

     847 .................AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT 879
                         ||||||||||||   |||:::      ||||||
     117 roAsnArgProArgLeuArgArgLysArgGlnValArg......ArgGly 131

     880 .........ATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAA 920
                 :::   :::|||   ||||||:::|||||||:::::::::||
     132 HisProThrValHisSerGluThrLysTyrValGluLeuIleValIleAs 148

     921 TGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAACA 970
         |||||||:::   :::::::::   |||:::|||   :::   |||:::|
     148 nAspHisGlnLeuPheGluGlnMetArgGlnSerValValLeuThrSerA 165

     971 ACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAG 1020
         ||||||||||||||||||||||||:::|||   ||||||||||||||||
     165 snPheAlaLysSerValValAsnLeuAlaAspValIleTyrLysGluGln 181

    1021 CTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGA 1070
         ||||||||||||::::||||||||||||:::|||||||||:::::   ||
     182 LeuAsnThrArgIleValLeuValAlaMetGluThrTrpAlaAspGlyAs 198

    1071 TCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAA 1120
         |:::|||::::::      :::   ::::::   |||   ::::::
```

```
 198 pLysIleGlnValGlnAspAspLeuLeuGluThrLeuAlaArgLeuMetV 215

1121 AATACCGGCAGCGC...ATTAAGCAGCATGCTGATGCTGTGCACCTCATC 1167
     ||||||:::::   :::  :::   ::::::|||:::|||||||:::
 215 alTyrArgArgGluGlyLeuProGluProSerAsnAlaThrHisLeuPhe 231

1168 TCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGG 1217
     |||      ||||||:::          |||:::   :::|||   |||||
 232 SerGlyArgThrPheGlnSerThrSerSerGlyAlaAlaTyrValGlyGl 248

1218 TGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAA 1267
     |:::||||||   :::::::|||   ||||||||||||||||||
 248 yIleCysSerLeuSerHisGlyGlyGlyValAsnGluTyrGlyAsnMetG 265

 268 TGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATC 1317
     |||:::|||   :::|||:::|||:::|||:::|||||||||||||:::
 265 lyAlaMetAlaValThrLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMet 281

1318 CAATGG...GAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATC 1364
     :::|||   :::      |||              |||  .|||   :::
 282 MetTrpAsnLysHisArgSerSerAlaGlyAspCysLysCysProAspIl 298

1365 CTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAAT 1414
      |||   |||||||||||||||:::|||||||           |||||||
 298 eTrpLeuGlyCysIleMetGluAspThrGlyPheTyrLeuProArgLysP 315

1415 TTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGT 1464
     |||||:::|||||||||   ||||||::::::|||||||:::|||||
 315 heSerArgCysSerIleAspGluTyrAsnGlnPheLeuGlnGluGlyGly 331

1465 GGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATG 1514
     |||:::|||||||||||||:::|||   ||||||:::::::|||   |||||
 332 GlySerCysLeuPheAsnLysProLeuLysLeuLeuAspProProGluCy 348

1515 TGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCAATG 1564
     |||||||||:::|||||||||||||||||||||||||||||||
 348 sGlyAsnGlyPheValGluAlaGlyGluGluCysAspCysGlySerValG 365

1565 TGGAATGCTAT.........GGATTATGCTGTAAGAAATGTTCCCTCTCC 1605
      ||||||             |||   |||||||||||||||:::|||:::
 365 lnGluCysSerArgAlaGlyGlyAsnCysCysLysLysCysThrLeuThr 381

1606 AACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCT 1655
     :::   |||   ||||||||||||   ||||||::::::      |||
 382 HisAspAlaMetCysSerAspGlyLeuCysCysArgArg......CysLy 396

 656 TTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATA 1705
      ::::::|||||||||   :::||||||:::||||||||||||||||||
 396 sTyrGluProArgGlyValSerCysArgGluAlaValAsnGluCysAspI 413

1706 TTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCAT 1755
     ||:::|||   |||||||||||||||:::|||||||||||||||||||
 413 leAlaGluThrCysThrGlyAspSerSerGlnCysProProAsnLeuHis 429

1756 AAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGG 1805
     |||   |||||||||   |||:::::::::|||||||||||||||:::||
 430 LysLeuAspGlyTyrTyrCysAspHisGluGlnGlyArgCysTyrGlyGl 446

1806 CGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGG 1855
     |:::|||||||||||||||:::|||||||||   :::||||||      |
 446 yArgCysLysThrArgAspArgGlnCysGlnValLeuTrpGlyHisAlaA 463

1856 CTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACT 1905
     |||||      |||:::|||||||||||||||||||||:::|||||||||
 463 laAla......AspArgPheCysTyrGluLysLeuAsnValGluGlyThr 477

1906 GAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAA 1955
     |||:::|||:::|||||||:::   |||:::   |||:::||||||||||
 478 GluArgGlySerCysGlyArgLysGlySerGlyTrpValGlnCysSerLy 494

1956 ACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTC 2005
     |:::|||||:::|||||||||||||||||||:::|||:::::   ||||
 494 sGlnAspValLeuCysGlyPheLeuLeuCysValAsnIleSerGlyAlaP 511

2006 CACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCAT 2055
     |||||:::|||:::|||   |||:::|||      ::::::||||||||
 511 roArgLeuGlyAspLeuValGlyAspIleSerSerValThrPheTyrHis 527

2056 CAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGA 2105
     ||||||:::   :::||||||  |||:::|||||   |||   |||
 528 GlnGlyLysGluLeuAspCysArgGlyGlyHisValGlnLeuAlaAspGl 544

2106 TACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGA 2155
      :::|||::::::|||||||||||||||||||   |||:|||||:::|||:
 544 ySerAspLeuSerTyrValGluAspGlyThrAlaCysGlyProAsnMetL 561

2156 TGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGC 2205
     ::|||||||||:::::|||||      :::|||:::|||:::|||:::
 561 euCysLeuAspHisArgCysLeuProAlaSerAlaPheAsnPheSerThr 577

2206 TGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAG 2255
     ||||||   :::::::::   ::::::|||||   ||||||||||||||
 578 CysProGlySerGlyGluArgArgIleCysSerHisHisGlyValCysSe 594

2256 TAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCA 2305
     |||||||:::   |||||||||:::      |||:::|||   |||||||
 594 rAsnGluGlyLysCysIleCysGlnProAspTrpThrGlyLysAspCysS 611

2306 GTATCCGGGATCCAGTTAGGAACCTTCACCCCCCC...AAGGATGAAGGA 2352
     |||||::::::|||:::   :::      |||      :::   |||
 611 erIleHisAsnProLeuProThrSerProProThrGlyGluThrGluArg 627

2353 CCCAAGGGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGGTGC 2402
        ||||||||||||:::|||||||:::|||||||||||||||||||||||
 628 TyrLysGlyProSerGlyThrAsnIleIleIleIleGlySerIleAlaGlyAl 644

2403 CATCCTGGTAGCAGCTATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAA 2452
     |:::|||||||||||||   ||||||||||||||||||||||||||||||
 644 aValLeuValAlaAlaTyrValLeuGlyGlyThrGlyTrpGlyPheLysA 661

2453 ATGTCAAGAAGAGAAGG 2469
     ||:::::::::   |||
 661 snIleArgArgGlyArg 666
```

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:AAR67758

seq_documentation_block:

```
ID   AAR67758 standard; Protein; 524 AA.
XX
AC   AAR67758;
XX
DT   01-AUG-1995  (first entry)
XX
DE   Human fetal brain MDC protein.
XX
KW   MDC protein; breast cancer; mamma carcinoma; ovary cancer;
KW   chromosome-17.
XX
OS   Homo sapiens.
XX
PN   EP633268-A.
XX
PD   11-JAN-1995.
XX
PF   13-MAY-1994;   94EP-0107487.
XX
PR   14-MAY-1993;   93JP-0136602.
PR   22-SEP-1993;   93JP-0257455.
PR   23-FEB-1994;   94JP-0049904.
PR   12-APR-1994;   94JP-0073328.
XX
PA   (CANC-) CANCER INST.
PA   (EISA ) EISAI CO LTD.
XX
PI   Emi M,  Nakamura Y;
XX
```

```
DR   WPI; 1995-038478/06.
DR   N-PSDB; AAQ76120.
XX
PT   Novel MDC protein and DNA encoding it - used to develop prods.
PT   for the study, diagnosis and therapy of cancers, partic. breast
PT   and ovarian cancer
XX
PS   Disclosure; Page 86-89; 123pp; English.
XX
CC   A detailed map of human chromosome-17 was constructed to analyze the
CC   chromosome in breast and ovarian cancer tissues.  2 Novel cDNAs
CC   (given in AAQ76120-21) were obtained from fetal cerebellum that
CC   encoded novel MDC proteins (AAR67758-59, respectively).  5'-RACE and
CC   RT-PCR revealed the sequence given in AAQ76122, encoding the MDC given
CC   in AAR75352.  Sequences common for these clones are given in AAQ76119
CC   and AAR66757, respectively.  A genomic DNA sequence (AAQ76124) for MDC
CC   was obtained from cosmid clone cCI17-904.
XX
SQ   Sequence   524 AA;
```

[redacted]ment_scores:
Quality: 1297.00 Length: 517
Ratio: 3.515 Gaps: 12
Percent Similarity: 71.373 Percent Identity: 49.516

alignment_block:
US-09-634-252A-2 x AAR67758 ..

Align seg 1/1 to: AAR67758 from: 1 to: 524

```
 502 CTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTAC...GA 548
     |||::::::   ||||||||||||:::||||||||||   |||:::   ::
   4 LeuSerHisGlnLeuLeuSerSerGlnTyrValGluArgHisPheSerAr 20

 549 AAATGGGAAACCACAGTACTCTAAGGGT...GGAGAGCACTGTTACTACC 595
     ::::|||      |||:::|||   |||   |||:::||||||||||||:
  20 gGluGlyThrThrGlnHisSerThrGlyAlaGlyAspHisCysTyrTyrG 37

 596 ATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGC 645
     ::|||::::::||||||         |||   :::|||||||||||||||
  37 lnGlyLysLeuArgGlyAsnProHisSerPheAlaAlaLeuSerThrCys 53

 646 AATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGA 695
     :::||||||||||||:::|||:::|||   :::::::::|||::::::||
  54 GlnGlyLeuHisGlyValPheSerAspGlyAsnLeuThrTyrIleValGl 70

 696 GCCACTAGAGCTGGTT......CATGATGAGAAAAGCACAGGTCGACCAC 739
     ||||   |||::::::                :::          ||||
  70 uProGlnGluValAlaGlyProTrpGlyAlaProGlnGlyProLeuProH 87

 740 ATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAAT 789
     ||:::|||   :::|||
  87 isLeuIleTyrArgThr..................................... 92

 790 CTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTG 839
                               |||:::|||   :::
  93 ..........................ProLeuLeuProAspProLeuGl 100

 840 GTTGAAA.......................................... 846
         :::
 100 yCysArgGluProGlyCysLeuPheAlaValProAlaGlnSerAlaProP 117

 847 ................AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT 879
                     ||||||||||||   |||:::      ||||||
 117 roAsnArgProArgLeuArgArgLysArgGlnValArg......ArgGly 131

 880 .........ATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAA 920
             :::   :::|||   ||||||:::||||||::::::::||
 132 HisProThrValHisSerGluThrLysTyrValGluLeuIleValIleAs 148

 921 TGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAACA 970
     |||||||:::   :::::::::   |||:::|||   :::   |||:::|
```

```
 148 nAspHisGlnLeuPheGluGlnMetArgGlnSerValValLeuThrSerA 165

 971 ACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAG 1020
     ||||||||||||||||||||||||||:::|||   ||||||||||||||
 165 snPheAlaLysSerValValAsnLeuAlaAspValIleTyrLysGluGln 181

1021 CTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGA 1070
     ||||||||||||:::||||||||||||:::|||||||||::::::   ||
 182 LeuAsnThrArgIleValLeuValAlaMetGluThrTrpAlaAspGlyAs 198

1071 TCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAA 1120
     |:::|||::::::      :::   ::::::   |||   ::::::
 198 pLysIleGlnValGlnAspAspLeuLeuGluThrLeuAlaArgLeuMetV 215

1121 AATACCGGCAGCGC...ATTAAGCAGCATGCTGATGCTGTGCACCTCATC 1167
       ||||||::::::   :::   :::   ::::::|||:::||||||:::
 215 alTyrArgArgGluGlyLeuProGluProSerAsnAlaThrHisLeuPhe 231

1168 TCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGG 1217
     |||      ||||||:::         |||:::   :::|||   |||||
 232 SerGlyArgThrPheGlnSerThrSerSerGlyAlaAlaTyrValGlyGl 248

1218 TGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAA 1267
     |:::||||||   :::::::|||   ||||||||||||||||||
 248 yIleCysSerLeuSerHisGlyGlyGlyValAsnGluTyrGlyAsnMetG 265

1268 TGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATC 1317
       |||:::|||   :::|||:::|||:::|||:::||||||||||||:::
 265 lyAlaMetAlaValThrLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMet 281

1318 CAATGG...GAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATC 1364
     :::|||   :::      |||           |||   |||   :::
 282 MetTrpAsnLysHisArgSerSerAlaGlyAspCysLysCysProAspIl 298

1365 CTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAAT 1414
      |||   ||||||||||||||:::||||||            |||||||
 298 eTrpLeuGlyCysIleMetGluAspThrGlyPheTyrLeuProArgLysP 315

1415 TTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGT 1464
     ||||:::||||||||   ||||||::::::||||||||:::||||||
 315 heSerArgCysSerIleAspGluTyrAsnGlnPheLeuGlnGluGlyGly 331

1465 GGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATG 1514
     |||:::|||||||||||||:::|||   ||||||::::::|||   |||||
 332 GlySerCysLeuPheAsnLysProLeuLysLeuLeuAspProProGluCy 348

1515 TGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATG 1564
     |||||||||:::|||||||||||||||||||||||||||||||||||||
 348 sGlyAsnGlyPheValGluAlaGlyGluGluCysAspCysGlySerValG 365

1565 TGGAATGCTAT.........GGATTATGCTGTAAGAAATGTTCCCTCTCC 1605
       ||||||           |||   |||||||||||||||:::|||:::
 365 lnGluCysSerArgAlaGlyGlyAsnCysCysLysLysCysThrLeuThr 381

1606 AACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCT 1655
     :::   |||   ||||||||||||   ||||||::::::      |||
 382 HisAspAlaMetCysSerAspGlyLeuCysCysArgArg......CysLy 396

1656 TTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATA 1705
      ::::::|||||||||   :::||||||:::||||||||||||||||||
 396 sTyrGluProArgGlyValSerCysArgGluAlaValAsnGluCysAspI 413

1706 TTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCAT 1755
     ||:::|||   ||||||||||||||:::||||||||||||||||||||
 413 leAlaGluThrCysThrGlyAspSerSerGlnCysProProAsnLeuHis 429

1756 AAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGG 1805
     |||   |||||||||   |||:::::::::||||||||||||||:::||
 430 LysLeuAspGlyTyrTyrCysAspHisGluGlnGlyArgCysTyrGlyGl 446

1806 CGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGG 1855
     |:::|||||||||||||||:::|||||||||   :::||||||      |
 446 yArgCysLysThrArgAspArgGlnCysGlnValLeuTrpGlyHisAlaA 463
```

```
    1856 CTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACT 1905
         |||||      |||:::|||||||||||||||||||||||:::|||||||||
     463 laAla......AspArgPheCysTyrGluLysLeuAsnValGluGlyThr 477

    1906 GAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAA 1955
         |||:::|||:::||||||:::   |||:::   |||:::||||||||||||
     478 GluArgGlySerCysGlyArgLysGlySerGlyTrpValGlnCysSerLy 494

    1956 A 1956
         |
     494 s 494

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT:AAR67757

[redacted]ocumentation_block:
AAR67757 standard; Protein; 488 AA.

XX
AC   AAR67757;
XX
DT   01-AUG-1995  (first entry)
XX
DE   Human fetal brain MDC protein.
XX
KW   MDC protein; breast cancer; mamma carcinoma; ovary cancer;
KW   chromosome-17.
XX
OS   Homo sapiens.
XX
PN   EP633268-A.
XX
PD   11-JAN-1995.
XX
PF   13-MAY-1994;   94EP-0107487.
XX
PR   14-MAY-1993;   93JP-0136602.
PR   22-SEP-1993;   93JP-0257455.
PR   23-FEB-1994;   94JP-0049904.
PR   12-APR-1994;   94JP-0073328.
XX
PA   (CANC-) CANCER INST.
PA   (EISA ) EISAI CO LTD.
XX
PI   Emi M,  Nakamura Y;
XX
     WPI; 1995-038478/06.
     N-PSDB; AAQ76119.

     Novel MDC protein and DNA encoding it - used to develop prods.
PT   for the study, diagnosis and therapy of cancers, partic. breast
PT   and ovarian cancer
XX
PS   Disclosure; Page 73-74; 123pp; English.
XX
CC   A detailed map of human chromosome-17 was constructed to analyze the
CC   chromosome in breast and ovarian cancer tissues.  2 Novel cDNAs
CC   (given in AAQ76120-21) were obtained from fetal cerebellum that
CC   encoded novel MDC proteins (AAR67758-59, respectively).  5'-RACE and
CC   RT-PCR revealed the sequence given in AAQ76122, encoding the MDC given
CC   in AAR75352.  Sequences common for these clones are given in AAQ76119
CC   and AAR66757, respectively.  A genomic DNA sequence (AAQ76124) for MDC
CC   was obtained from cosmid clone cCI17-904.
XX
SQ   Sequence   488 AA;


alignment_scores:
            Quality: 1293.00              Length:    513
              Ratio:  3.533                Gaps:     12
 Percent Similarity:  71.345   Percent Identity: 49.708

alignment_block:
 US-09-634-252A-2 x AAR67757   ..
```

```
Align seg 1/1  to: AAR67757  from: 1  to: 488

    514 TTGTTGTCTTCTGATTATGTGGAGATTCACTAC...GAAAATGGGAAACC 560
        ||||||||||||:::|||||||||   |||:::   ::::::|||
      1 LeuLeuSerSerGlnTyrValGluArgHisPheSerArgGluGlyThrTh 17

    561 ACAGTACTCTAAGGGT...GGAGAGCACTGTTACTACCATGGAAGCATCA 607
         |||:::|||   |||   |||:::|||||||||||||:::|||::::::|
     17 rGlnHisSerThrGlyAlaGlyAspHisCysTyrTyrGlnGlyLysLeuA 34

    608 GAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCAT 657
        |||||          |||   :::|||||||||||||||:::|||||||||
     34 rgGlyAsnProHisSerPheAlaAlaLeuSerThrCysGlnGlyLeuHis 50

    658 GGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCT 707
        |||:::|||:::|||   :::::::::|||:::::|||||   |||::
     51 GlyValPheSerAspGlyAsnLeuThrTyrIleValGluProGlnGluVa 67

    708 GGTT......CATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGA 751
        ::::                  :::         ||||||:::|||   :
     67 lAlaGlyProTrpGlyAlaProGlnGlyProLeuProHisLeuIleTyrA 84

    752 AAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAA 801
        ::|||
     84 rgThr..................................................... 85

    802 AGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAA..... 846
                      |||:::|||   :::            :::
     86 ..............ProLeuLeuProAspProLeuGlyCysArgGluPr 97

    846 ..................................................... 846

     97 oGlyCysLeuPheAlaValProAlaGlnSerAlaProProAsnArgProA 114

    847 .....AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT.........ATA 882
             ||||||||||||   |||:::      ||||||          :::
    114 rgLeuArgArgLysArgGlnValArg......ArgGlyHisProThrVal 128

    883 TTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAAC 932
           :::|||   ||||||:::|||||||:::::::::||||||||||:::
    129 HisSerGluThrLysTyrValGluLeuIleValIleAsnAspHisGlnLe 145

    933 GTATAAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGT 982
         :::::::::   |||:::|||   :::   |||:::||||||||||||||
    145 uPheGluGlnMetArgGlnSerValValLeuThrSerAsnPheAlaLysS 162

    983 CCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGG 1032
        ||||||||||||||:::|||   |||||||||||||||||||||||||||
    162 erValValAsnLeuAlaAspValIleTyrLysGluGlnLeuAsnThrArg 178

   1033 GTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACAT 1082
        :::|||||||||||||:::|||||||||::::::   |||:::|||:::::
    179 IleValLeuValAlaMetGluThrTrpAlaAspGlyAspLysIleGlnVa 195

   1083 CACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGC 1132
        :      :::   ::::::   |||   ::::::      ||||||:::::
    195 lGlnAspAspLeuLeuGluThrLeuAlaArgLeuMetValTyrArgArgG 212

   1133 GC...ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACA 1179
        ::   :::   :::   ::::::|||:::||||||:::|||      |||
    212 luGlyLeuProGluProSerAsnAlaThrHisLeuPheSerGlyArgThr 228

   1180 TTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCG 1229
        |||:::         |||:::   :::|||   |||||||:::||||||
    229 PheGlnSerThrSerSerGlyAlaAlaTyrValGlyGlyIleCysSerLe 245

   1230 CACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCAC 1279
         ::::::|||   ||||||||||||||||         |||:::|||
    245 uSerHisGlyGlyGlyValAsnGluTyrGlyAsnMetGlyAlaMetAlaV 262

   1280 AAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGG...GAA 1326
          :::|||:::|||:::|||:::||||||||||||:::::|||   :::
```

```
 262 alThrLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMetMetTrpAsnLys 278

1327 CCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTG 1376
          |||          |||   |||   :::   |||   |||||
 279 HisArgSerSerAlaGlyAspCysLysCysProAspIleTrpLeuGlyCy 295

1377 CATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCA 1426
     ||||||||||:::|||||||           |||||||||||||:::||||
 295 sIleMetGluAspThrGlyPheTyrLeuProArgLysPheSerArgCysS 312

1427 GCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTT 1476
     |||||   ||||||:::::|||||||||:::|||||||||:::||||||
 312 erIleAspGluTyrAsnGlnPheLeuGlnGluGlyGlyGlySerCysLeu 328

1477 TTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATA 1526
     ||||||:::|||   ||||||::::::|||   ||||||||||||||::
 329 PheAsnLysProLeuLysLeuLeuAspProProGluCysGlyAsnGlyPh 345

1527 CGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTAT. 1575
     :|||||||||||||||||||||||||||||          ||||||
 345 eValGluAlaGlyGluGluCysAspCysGlySerValGlnGluCysSerA 362

 576 ........GGATTATGCTGTAAGAAATGTTCCCTCTCCAACGGGGCTCAC 1617
             |||   |||||||||||||||:::|||::::::   |||
 362 rgAlaGlyGlyAsnCysCysLysLysCysThrLeuThrHisAspAlaMet 378

1618 TGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACG 1667
     ||||||||||||   ||||||::::::      |||   ::::::|||||
 379 CysSerAspGlyLeuCysCysArgArg......CysLysTyrGluProAr 393

1668 AGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATT 1717
     ||||   :::|||||||:::|||||||||||||||||||||:::|||   |
 393 gGlyValSerCysArgGluAlaValAsnGluCysAspIleAlaGluThrC 410

1718 GTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGA 1767
     ||||||||||||||:::|||||||||||||||||||||||||   ||||||
 410 ysThrGlyAspSerSerGlnCysProProAsnLeuHisLysLeuAspGly 426

1768 TATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGAC 1817
     |||   |||:::::::::|||||||||||||||:::|||:::|||||||
 427 TyrTyrCysAspHisGluGlnGlyArgCysTyrGlyGlyArgCysLysTh 443

1818 CAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTG 1867
     |||||||:::|||||||||   :::||||||      ||||||      |
 443 rArgAspArgGlnCysGlnValLeuTrpGlyHisAlaAlaAla......A 458

1868 ACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAAC 1917
     ||:::|||||||||||||||||||||:::|||||||||||||:::|||:::
 458 spArgPheCysTyrGluLysLeuAsnValGluGlyThrGluArgGlySer 474

1918 TGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAA 1956
     ||||||:::   |||:::   |||:::|||||||||||
 475 CysGlyArgLysGlySerGlyTrpValGlnCysSerLys 487

     ame: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB50942

seq_documentation_block:
ID   AAB50942 standard; peptide; 778 AA.
XX
AC   AAB50942;
XX
DT   20-MAR-2001  (first entry)
XX
DE   ADAM gene #1 peptide #1.
XX
KW   ADAM; disintegrin; metalloprotease; food additive; breast cancer;
KW   ovarian.
XX
OS   Homo sapiens.
XX
PN   WO200073323-A2.
XX
PD   07-DEC-2000.
XX
PF   25-MAY-2000; 2000WO-US14308.
XX
PR   27-MAY-1999;   99US-0136388.
PR   09-JUL-1999;   99US-0142930.
PR   28-JAN-2000; 2000US-0178717.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Wei P,  Ni J,  Hastings GA,  Shi Y;
XX
DR   WPI; 2001-016507/02.
XX
PT   Seven nucleic acid molecules encoding ADAM polypeptides containing a
PT   disintegrin and metalloprotease domain, useful in the prevention,
PT   treatment and diagnosis of cancer, immune disorders, cardiovascular
PT   disorders and neurological diseases -
XX
PS   Disclosure; Page 12; 287pp; English.
XX
CC   The present invention relates to seven members of the ADAMs (proteins
CC   which contain A Disintegrin And Metalloprotease domain) protein family.
CC   The ADAMs proteins and DNA may be used to treat disease, as a food
CC   additive or preservative, for chromosome identification, as probes
CC   for diagnosing a disorder related to the female reproductive system,
CC   particularly breast and/or ovary cancer. They are also useful in the
CC   gene therapy of breast and ovarian cancer.
XX
SQ   Sequence   778 AA;


alignment_scores:
           Quality: 1007.50              Length:    801
             Ratio:  2.099                 Gaps:     29
 Percent Similarity:  59.925   Percent Identity: 31.461

alignment_block:
 US-09-634-252A-2 x AAB50942   ..

 Align seg 1/1  to: AAB50942  from: 1  to: 778

 112 CCGGCCCGCACGCCGCCCTGCCGCCTGCTTCTCGTCCTTCTCCTGCTGCC 161
     |||:::   :::      :::||||||   |||:::   :::   |||
   2 ProGlyGlyAlaGlyAlaAlaArgLeuCysLeuLeuAlaPheAlaLeuGl 18

 162 TCCGCTCGCCGCCTCGTCCCGGCCCCGCGCCTGGGGGGCTGCTGCGCCCA 211
      ||||||            |||||||||               |||
  18 nProLeu............ArgProArg.............AlaAlaA 26

 212 GCGCTCCGCATTGGAATGAAACTGCAGAAAAAAATTTGGGAGTCCTGGCA 261
          |||   |||::::::   :::|||::::::
  26 rgGluProGlyTrpThrArgGlySerGluGluGly.............. 37

 262 GATGAAGACAATACATTGCAACAGAATAGCAGCAGTAATATCAGTTACAG 311
                                                      ||
  38 ..............................................Se 38

 312 CAATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACA 361
     |      :::|||   |||:::   :::|||:::
  38 rProLysLeuGlnHisGluLeuIleIleProGlnTrp............. 50

 362 TCAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGA 411
         :::   ||||||||||||
  51 .....LysThrSerGluSerProVal..................... 57

 412 CACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCA 461
                       :::      |||      :::|||::::::::
  58 .................ArgGluLysHisProLeuLysAlaGluLeuAr 68

 462 GATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATG 511
     ::::   |||   |||   ::::::||||||||||||      .|||:::
  68 gValMetAlaGluGlyArgGluLeuIleLeuAspLeuGluLysAsnGluG 85
```



```
     655 lyLysLysCysAsnGlyHisGlyValCysAsnAsnAsnGlnAsnCysHis 671

    2275 TGTGATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGTTAG 2324
         |||         ||||||         |||:::
     672 CysLeuProGlyTrpAlaProProPheCysAsnThrProGlyHisGlyGl 688

    2325 GAACCTTCACCCCCCCAAGGATGAAGGACCCAAGGGTCCTAGTGCCACCA 2374
          ::::::            |||:::||||||      |||:::::::::::
     688 ySerIle............AspSerGlyProMetProProGluSerValG 701

    2375 ATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCTATTGTC 2424
         ::   ::::::::::::::      ||||||||||||   |||::::::
     701 lyProValValAlaGlyValLeuValAlaIleLeuValLeuAlaValLeu 717

    2425 CTT 2427
         :::
     718 Met 718

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB50935

documentation_block:
AAB50935 standard; protein; 918 AA.

AAB50935;
XX
DT   20-MAR-2001  (first entry)
XX
DE   ADAM protein #1.
XX
KW   ADAM; disintegrin; metalloprotease; food additive; breast cancer;
KW   ovarian.
XX
OS   Homo sapiens.
XX
PN   WO200073323-A2.
XX
PD   07-DEC-2000.
XX
PF   25-MAY-2000; 2000WO-US14308.
XX
PR   27-MAY-1999;   99US-0136388.
PR   09-JUL-1999;   99US-0142930.
PR   28-JAN-2000; 2000US-0178717.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Ruben SM,  Wei P,  Ni J,  Hastings GA,  Shi Y;
XX
DR   WPI; 2001-016507/02.
XX
PT   Seven nucleic acid molecules encoding ADAM polypeptides containing a
PT   disintegrin and metalloprotease domain, useful in the prevention,
PT   treatment and diagnosis of cancer, immune disorders, cardiovascular
PT   disorders and neurological diseases -
XX
PS   Claim 11; Page 271-274; 287pp; English.
XX
CC   The present invention relates to seven members of the ADAMs (proteins
CC   which contain A Disintegrin And Metalloprotease domain) protein family.
CC   The ADAMs proteins and DNA may be used to treat disease, as a food
CC   additive or preservative, for chromosome identification, as probes
CC   for diagnosing a disorder related to the female reproductive system,
CC   particularly breast and/or ovary cancer. They are also useful in the
CC   gene therapy of breast and ovarian cancer.
XX
SQ   Sequence   918 AA;


alignment_scores:
           Quality: 1007.50            Length:    801
             Ratio:  2.099              Gaps:     29
 Percent Similarity:  59.925   Percent Identity: 31.461

alignment_block:
 US-09-634-252A-2 x AAB50935   ..

 Align seg 1/1  to: AAB50935  from: 1  to: 918

     112 CCGGCCCGCACGCCGCCCTGCCGCCTGCTTCTCGTCCTTCTCCTGCTGCC 161
         |||:::   :::      :::||||||   |||:::   :::   |||
       2 ProGlyGlyAlaGlyAlaAlaArgLeuCysLeuLeuAlaPheAlaLeuGl 18

     162 TCCGCTCGCCGCCTCGTCCCGGCCCCGCGCCTGGGGGGCTGCTGCGCCCA 211
          ||||||            |||||||||               |||
      18 nProLeu............ArgProArg..............AlaAlaA 26

     212 GCGCTCCGCATTGGAATGAAACTGCAGAAAAAAATTTGGGAGTCCTGGCA 261
             |||   |||::::::   :::|||::::::
      26 rgGluProGlyTrpThrArgGlySerGluGluGly.............. 37

     262 GATGAAGACAATACATTGCAACAGAATAGCAGCAGTAATATCAGTTACAG 311
                                                        ||
      38 ..............................................Se 38

     312 CAATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACA 361
         |      :::|||   |||:::   :::|||:::
      38 rProLysLeuGlnHisGluLeuIleIleProGlnTrp............. 50

     362 TCAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGA 411
             :::   ||||||||||||
      51 .....LysThrSerGluSerProVal....................... 57

     412 CACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCA 461
                         :::      |||      :::|||:::::::
      58 .................ArgGluLysHisProLeuLysAlaGluLeuAr 68

     462 GATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATG 511
         ::::   |||   |||   :::::||||||||||||      |||:::
      68 gValMetAlaGluGlyArgGluLeuIleLeuAspLeuGluLysAsnGluG 85

     512 GTTTGTTGTCTTCTGATTATGTGGAGATTCACTAC...GAAAATGGGAAA 558
           |||::::::   :::|||:::|||   ||||||   ::::::|||:::
      85 lnLeuPheAlaProSerTyrThrGluThrHisTyrThrSerSerGlyAsn 101

     559 CCACAGTACTCT....AAGGGTGGAGAGCACTGTTACTACCATGGAAG 602
         ||||||   :::      |||      :::||||||:::|||||||||::
     102 ProGlnThrThrThrArgLysLeuGluAspHisCysPheTyrHisGlyTh 118

     603 CATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGAC 652
         ::::|||   ::::::   |||:::|||:::||||||||||||:::|||:
     118 rValArgGluThrGluLeuSerSerValThrLeuSerThrCysArgGlyI 135

     653 TTCATGGCATGTTTGAA...GATGATACCTTCGTGTATATGATAGAGCCA 699
         :::::|||::::::      ::::::::::::   |||:::|||||||||
     135 leArgGlyLeuIleThrValSerSerAsnLeuSerTyrValIleGluPro 151

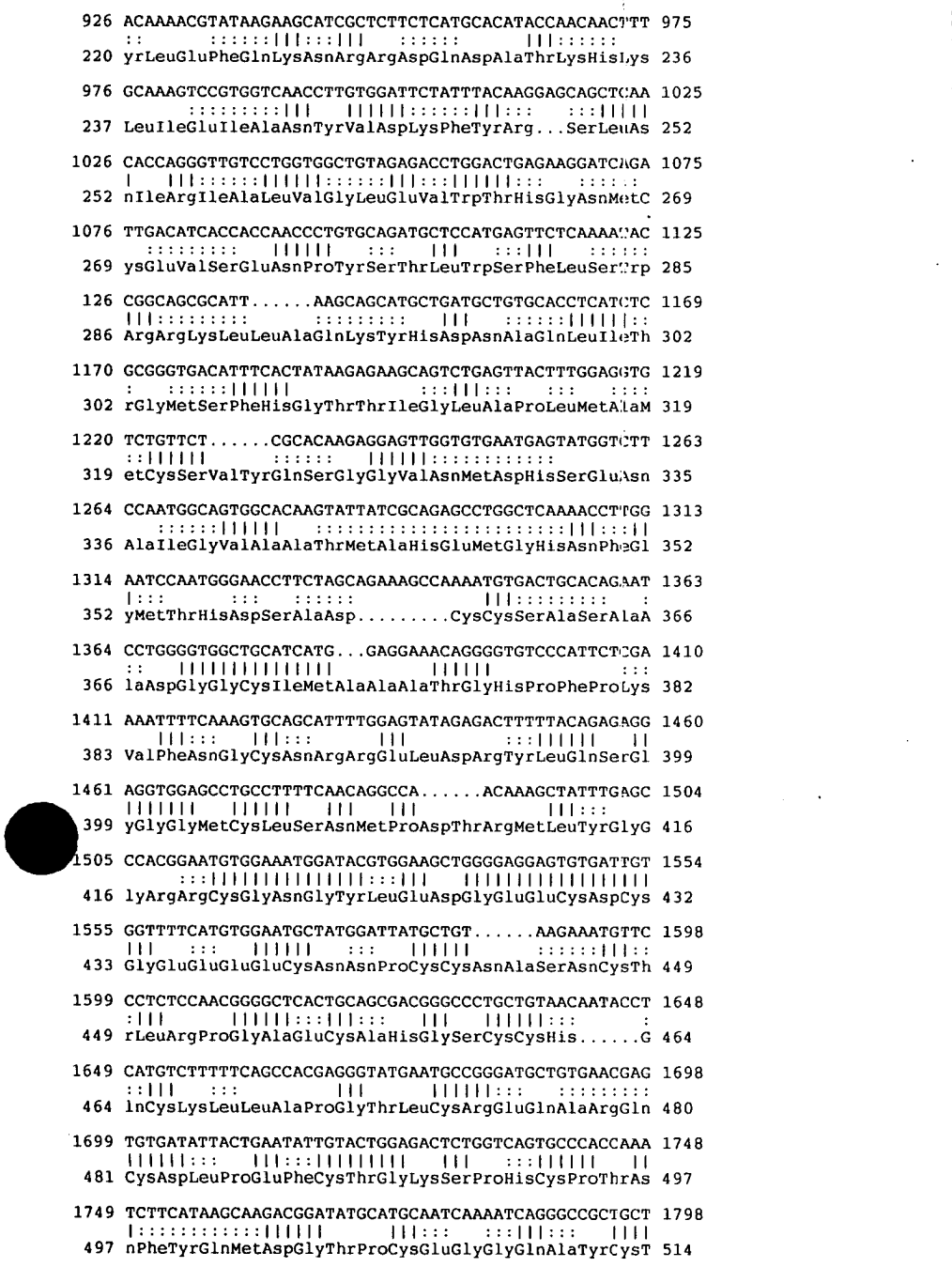
     700 CTAGAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCACATATAATCCA 749
         |||              |||:::|||:::          |||:::|||
     152 LeuPro.........AspSerLysGly.........GlnHisLeuIleTy 162

     750 GAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAG............A 787
          :::                 |||::::::::::|||            |
     162 rArg.................SerGluHisLeuLysProProProGlyA 173

     788 ATCTCACTATGGAAAGAGGTGAC............CAGTGGCCCTTTCTC 825
         ||      :::|||:::::                :::|||      |||
     173 snCysGlyPheGluHisSerLysProThrThrArgAspTrpAla...Leu 188

     826 TCTGAATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACG 875
         :::      |||      |||:::|||   |||   :::      :::||
     189 GlnPheThrGlnGlnThrLysLysArgProArgArgMet......LysAr 203

     876 TGGTATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAATGATC 925
         |      :::::::::|||||||||:::||||||   :::|||   |||:
     203 gGluAspLeuAsnSerMetLysTyrValGluLeuTyrLeuValAlaAspT 220
```



seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW25716

seq_documentation_block:
ID AAW25716 standard; Protein; 920 AA.
XX
AC AAW25716;
XX
DT 08-MAY-1998 (first entry)
XX
DE Mouse beta meltrin protein.
XX
KW Meltrin; myotube formation; muscle tissue; therapy; membrane protein;
KW myoblast; adhesion; fusion; medicinal; murine; mouse.
XX
OS Mus sp.
XX
PN WO9731109-A1.
XX

PD 28-AUG-1997.
XX
PF 17-OCT-1996; 96WO-JP03017.
XX
PR 23-FEB-1996; 96JP-0061756.
XX
PA (MOCH) MOCHIDA PHARM CO LTD.
XX
PI Fujisawa A, Mizushima C, Ogawa N, Shirakawa K, Yamakawa T;
XX
DR WPI; 1997-435161/40.
DR N-PSDB; AAT86008.
XX
PT Meltrin proteins involved with myotube formation in muscle - contain
PT metallo:proteinase and disintegrin domains
XX
PS Claim 4; Fig 3; 262pp; Japanese.
XX
CC This sequence represents a new membrane protein, beta meltrin, which
CC participates in the adhesion and fusion of myoblasts in the course of
CC myotube formation in muscle tissue. Meltrins and their partial peptides
CC can be used in the study of myotube formation, and in medicinal
CC compositions for muscle disorders. Meltrins contain an intracellular
CC domain, a membrane-penetrating domain, a fusion peptide domain, a
CC disintegrin domain, a metalloproteinase domain, a precursor domain and a
CC cysteine rich domain although these are not shown in the specification.
CC Meltrin antagonists can be used as bone absorption inhibitors, e.g. for
CC the treatment of osteoporosis and hypercalcaemia and for prevention of
CC cancer cell infiltration.
XX
SQ Sequence 920 AA;

alignment_scores:
Quality: 975.00 Length: 800
Ratio: 2.092 Gaps: 25
Percent Similarity: 58.250 Percent Identity: 30.500

alignment_block:
US-09-634-252A-2 x AAW25716 ..

Align seg 1/1 to: AAW25716 from: 1 to: 920

```
130 TGCCGCCTGCTTCTCGTCCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTC 179
    |||  |||  |||:::|||  |||     ||||||||||||  ::
 11 CysLeuLeuAlaLeuAlaLeuGlnLeuHisTrpProLeuAlaAlaCysGl 27

180 CCGGCCCCGCGCCTGGGGGGCTGCTGCGCCCAGCGCTCCGCATTGGAATG 229
    :                                  |||  |||:::
 27 u.................................ProGlyTrpThrT 32

230 AAACTGCAGAAAAAAATTTGGGAGTCCTGGCAGATGAAGACAATACATTG 279
       :::::::::
 32 hrArgGlySerGln..................................... 36

280 CAACAGAATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGA 329
                          :::  |||     :::|||  ||
 37 ......................GluGlySerProProLeuGlnHisGl.45

330 AATCACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGAAA 379
    |:::  :::|||:::                  :::  |||||||
 45 uLeuIleIleProGlnTrp.................ArgThrSerGluS 56

380 GCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAAT 429
    |||||                                    :::
 56 erPro.........................................Gly 58

430 AAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTC 479
    ::::::  |||     ;::|||:::::::::::::  |||  |||
 59 ArgGlyLysHisProLeuArgAlaGluLeuArgValMetAlaGluGlyAr 75

480 CAAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATT 529
     ::::::||||||||||||      |||:::  |||::::::      |
```

```
 75 gGluLeuIleLeuAspLeuGluLysAsnGluHisLeuPheAlaProAlaT 92

530 ATGTGGAGATTCACTAC...GAAAATGGGAAACCACAGTACTCT..... 570
    ||:::|||      |||      :::|||:::||||||  |||
 92 yrThrGluThrCysTyrThrAlaSerGlyAsnProGlnThrSerThrLeu 108

571 AAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGA 620
    |||:::  :::||||||:::|||||||||:::::|||  |||  ::
109 LysSerGluAspHisCysPheTyrHisGlyThrValArgAspValAspGl 125

621 CTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATG...TTTG 667
    :|||:::|||:::||||||||||||:::|||:::::|||:::  :::
125 uSerSerValThrLeuSerThrCysArgGlyIleArgGlyLeuIleIleV 142

668 AAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATGAT 717
       :::::::::  |||:::|||||||||:::
142 alArgSerAsnLeuSerTyrIleIleGluProVal.............. 153

718 GAGAAAAGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCAGGACA 767
       ::::::        |||  |||  ::::::
154 ...ProAsnSerAspSerGlnHisArgIleTyrArgSer........... 165

768 GTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAGGTGAC....... 810
                  ::::::||||||:::      |||:::
166 ..............GluHisLeuThrLeuProProGlyAsnCysGlyP 177

811 ........................CAGTGGCCCTTTCTCTCTGAATTA 834
                             :::|||  :::
177 heGluHisSerGlyProThrSerLysAspTrpAlaLeuGlnPheThrHis 193

835 CAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATT 884
    |||      |||:::  :::|||           :::|||      ::
194 GlnThrLysLysGlnProArgArgMet.........LysArgGluAspLe 207

885 TGAAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAACGT 934
    :::::::||||||||||:::||||||  :::|||  |||:::      :
207 uHisSerMetLysTyrValGluLeuTyrLeuValAlaAspTyrAlaGluP 224

935 ATAAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCC 984
    :::::|||:::|||  ::::::      |||::::::        :::
224 heGlnLysAsnArgHisAspGlnAspAlaThrLysArgLysLeuMetGlu 240

985 GTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGT 1034
    ::::::|||  ||||||:::::|||:::  :::||||||  |||::
241 IleAlaAsnTyrValAspLysPheTyrArg...SerLeuAsnIleArgIl 256

1035 TGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCA 1084
     ::::||||||:::::::|||:::||||||:::  |||:::  :::::::
256 eAlaLeuValGlyLeuGluValTrpThrHisGlyAspLysCysGluValS 273

1085 CCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGC 1134
     ::  ||||||  :::  |||  :::|||  ::::::|||::::::
273 erGluAsnProTyrSerThrLeuTrpSerPheLeuSerTrpArgArgLys 289

1135 ATT.........AAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGT 1175
     :::         |||:::|||  |||  ::::::||||||:::
290 LeuLeuAlaGlnLysSerHis...AspAsnAlaGlnLeuIleThrGlyAr 305

1176 GACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTT 1225
      :::|||:::            :::|||:::  :::  ::::::||||
305 gSerPheGlnGlyThrThrIleGlyLeuAlaProLeuMetAlaMetCysS 322

1226 CT......CGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATG 1269
     ||      ::::::  ||||||:::::::::::::          :::
322 erValTyrGlnSerGlyGlyValSerMetAspHisSerGluAsnAlaIle 338

1270 GCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCA 1319
     :::||||||:::::::::::::::::::::::::::|||:::|||:::::
339 GlyValAlaSerThrValAlaHisGluIleGlyHisAsnPheGlyMetSe 355

1320 ATGGGAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGG 1369
     :  :::  :::              |||:::::::::  :::  |
355 rHisAspSerAlaHis.........CysCysSerAlaSerAlaAlaAspG 369
```

```
1370 GTGGCTGCATCATG...GAGGAAACAGGGGTGTCCCATTCTCGAAAATTT 1416
     ||||||||||||||         ||||||             :::   |||
 369 lyGlyCysIleMetAlaAlaAlaThrGlyHisProPheProLysValPhe 385

1417 TCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGG 1466
     |||   |||:::      |||         :::||||||   ||||||||
 386 SerTrpCysAsnArgLysGluLeuAspArgTyrLeuGlnThrGlyGlyGl 402

1467 AGCCTGCCTTTTCAACAGGCCA......ACAAAGCTATTTGAGCCCACGG 1510
     |   ||||||   |||   |||              |||:::       :
 402 yMetCysLeuSerAsnMetProAspThrArgThrLeuTyrGlyGlyArgA 419

1511 AATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTGTT 1560
     ::|||||||||||||||:::|||   ||||||||||||||||||||||
 419 rgCysGlyAsnGlyTyrLeuGluAspGlyGluGluCysAspCysGlyGlu 435

1561 CATGTGGAATGCTATGGATTATGCTGT......AAGAAATGTTCCCTCTC 1604
     :::   ||||||   :::   ||||||      :::::::|||:::|| ::
 436 GluGluGluCysLysAsnProCysCysAsnAlaSerAsnCysThrLeuLy 452

1605 CAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTC 1654
     ::::||||||:::|||:::   |||   |||||||:::      :::|||
 452 sGluGlyAlaGluCysAlaHisGlySerCysCysHis......GlnCysL 467

1655 TTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGAT 1704
       :::         |||   :::||||||:::   |||::::::||||||
 467 ysLeuValAlaProGlyThrGlnCysArgGluGlnValArgGlnCysAsp 483

1705 ATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCA 1754
     :::   |||:::|||||||||   |||   :::||||||   |||   ::
 484 LeuProGluPheCysThrGlyLysSerProHisCysProThrAsnTyrTy 500

1755 TAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATG 1804
     :::::::||||||      |||:::   :::|||:::   ||||||||||
 500 rGlnMetAspGlyThrProCysGluGlyGlyGlnAlaTyrCysTyrAsnG 517

1805 GCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAG 1854
     ||   |||   |||   ::::::|||||||||   :::||||||
 517 lyMetCysLeuThrTyrGlnGluGlnCysGlnGlnLeuTrpGlyProGly 533

1855 GCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCAC 1904
     |||      :::      :::|||:::|||::::::|||:::   |||
 534 AlaArgProAlaLeuAspLeuCysPheGluArgValAsnAlaAlaGlyAs 550

1905 TGAGAAGGGAAACTGCGGGAAGGAT...GGAGACCGGTGGATTCAGTGCA 1951
           ||||||||||||||      :::   ::::::   :::||||
 550 pThrTyrGlyAsnCysGlyLysGlyLeuAsnGlyGlnTyrArgLysCysS 567

1952 GCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGA 2001
     ||   :::|||:::   ||||||   :::   |||   :::
 567 erProArgAspAlaLysCysGlyLysIleGlnCysGlnSerThrGlnAla 583

2002 GCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTA 2051
       |||         :::   ::::::      |||   |||::::::
 584 ArgProLeu......GluSerAsnAlaValSerIleAspThrThrIleTh 598

2052 CCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTA.... 2097
         :::|||||   |||   |||   |||:::||||||
 598 rLeuAsnGlyArgArgIleHisCysArgGlyThrHisValTyrArgGlyP 615

2098 ...................GATGATGATACGGATGTGGGCTATGTAGAA 2127
                         |||       |||   |||   |||
 615 roGluGluGluGluGlyGluGlyAspMetLeuAspProGlyLeuValMet 631

2128 GATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCT 2177
       ||||||   ||||||   :::   :::|||::::::   :::|||
 632 ThrGlyThrLysCysGlyHisAsnHisIleCysPheGluGlyGlnCysAr 648

2178 ACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCCAAGGGTA 2227
      :::   :::   ::::::   :::::::|||               ||||
 648 gAsnThrSerPheGluThrGluGlyCys..........GlyL 660
```

```
2228 AAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGT 2277
     ||   |||:::|||||||||||||||:::|||:::   :::|||   |||
 660 ysLysCysAsnGlyHisGlyValCysAsnAsnAsnLysAsnCysHisCys 676

2278 GATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGTTAGGAA 2327
               |||:::         |||:::                  ::
 677 PheProGlyTrpSerProProPheCysAsnThrProGlyAspGlyGlySe 693

2328 CCTTCACCCCCCCAAGGATGAAGGACCCAAGGGTCCTAGTGCCACCAATC 2377
     ::::             |||:::||||||      |||:::::::::::::
 693 rVal............AspSerGlyProLeuProProLysSerValGlyP 706

2378 TCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCTATTGTCCTT 2427
       :::|||:::::::::   :::|||:::::::|||   |||:::::::::
 706 roValIleAlaGlyValPheSerAlaLeuPheValLeuAlaValLeuVal 722
```

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW25715

seq_documentation_block:

```
ID   AAW25715 standard; Protein; 903 AA.
XX
AC   AAW25715;
XX
DT   08-MAY-1998  (first entry)
XX
DE   Mouse alpha meltrin protein.
XX
KW   Meltrin; myotube formation; muscle tissue; therapy; membrane protein;
KW   myoblast; adhesion; fusion; medicinal; murine; mouse.
XX
OS   Mus sp.
XX
PN   WO9731109-A1.
XX
PD   28-AUG-1997.
XX
PF   17-OCT-1996;   96WO-JP03017.
XX
PR   23-FEB-1996;   96JP-0061756.
XX
PA   (MOCH ) MOCHIDA PHARM CO LTD.
XX
PI   Fujisawa A,  Mizushima C,  Ogawa N,  Shirakawa K,  Yamakawa T;
XX
DR   WPI; 1997-435161/40.
DR   N-PSDB; AAT86007.
XX
PT   Meltrin proteins involved with myotube formation in muscle - contain
PT   metallo:proteinase and disintegrin domains
XX
PS   Claim 4; Fig 2; 262pp; Japanese.
XX
CC   This sequence represents a new membrane protein, alpha meltrin, which
CC   participates in the adhesion and fusion of myoblasts in the course of
CC   myotube formation in muscle tissue. Meltrins and their partial peptides
CC   can be used in the study of myotube formation, and in medicinal
CC   compositions for muscle disorders. Meltrins contain an intracellular
CC   domain, a membrane-penetrating domain, a fusion peptide domain, a
CC   disintegrin domain, a metalloproteinase domain, a precursor domain and a
CC   cysteine rich domain although these are not shown in the specification.
CC   Meltrin antagonists can be used as bone absorption inhibitors, e.g. for
CC   the treatment of osteoporosis and hypercalcaemia and for prevention of
CC   cancer cell infiltration.
XX
SQ   Sequence   903 AA;
```

```
alignment_scores:
           Quality:  966.00              Length:    822
             Ratio:  2.055                 Gaps:     32
 Percent Similarity:  57.178   Percent Identity: 31.630

alignment_block:
```

```
US-09-634-252A-2 x AAW25715   ..

Align seg 1/1  to: AAW25715  from: 1  to: 903

    103 GCCAGCGCCCCGGCC...CGCACGCCGCCCTGCCGCCTGCTTCTCGTCCT 149
        |||:::   ||||||   |||:::||||||:::|||   ||||||:::
      2 AlaGluArgProAlaArgArgAlaProProAlaArgAlaLeuLeuLeuAl 18

    150 TCTC.........CTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCGCG 190
         |||         ||||||   |||   ||||||   :::
     18 aLeuAlaGlyAlaLeuLeuAlaProArgAlaAlaArgGlyMet...SerL 34

    191 CCTGGGGGGCTGCTGCGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAA 240
          |||
     34 euTrp..................................................... 35

    241 AAAAATTTGGGAGTCCTGGCAGATGAAGACAATACATTGCAACAGAATAG 290
                                                :::|||:::::
     36 .....................................AspGlnArgGl 39

    291 CAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAATCACACTGC 340
        ::::   :::::::::   ::::::   ::::::|||:::      :::|
     39 yAlaTyrGluValAlaArgAlaSerLeuLeuSerLysAspProGlyIleP 56

    341 CTTCAAGACTCATATATTACATCAACCAAGACTCGGAAAGCCCTTATCAC 390
        ||:::::::   |||
     56 roGlyGlnSerIle..................................... 60

    391 GTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCA 440
                      |||:::                           ||
     61 ............ProAlaLysAsp.......................Hi 65

    441 TCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTC 490
        |            :::   |||::::::      :::      :::||||
     65 sProAspValLeuThrValGlnLeuGlnLeuGluSerArgAspLeuIleL 82

    491 TTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATT 540
        ||:::|||      |||:::||||||:::::::::   :::::::|||
     82 euSerLeuGluArgAsnGluGlyLeuIleAlaAsnGlyPheThrGluThr 98

    541 CACTACGAAAATGGGAAACCACAGTACTCT.........AAGGGTGGAGA 581
        ||||||   :::          :::   |||          :::      ::
     99 HisTyrLeuGlnAspGlyThrAspValSerLeuThrArgAsnHisThrAs 115

    582 GCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGG 631
        :|||||||||||||||||   ::::::|||          |||   |||:
    115 pHisCysTyrTyrHisGlyHisValGlnGlyAspAlaAlaSerValValS 132

    632 CTCTGTCAACCTGCAATGGACTTCATGGC......ATGTTTGAAGATGAT 675
        ::|||||||||||||:::   |||:::|||      ||||||||||:::
    132 erLeuSerThrCysSerAspLeuArgGlyLeuIleMetPheGluAsnLys 148

    676 ACCTTCGTGTATATGATAGAGCCACTA.....................GA 704
        |||      |||   :::||||||:::                     ::
    149 Thr......TyrSerLeuGluProMetLysAsnThrThrAspSerTyrLy 163

    705 GCTGGTTCATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAA 754
        :||||||      |||:::   |||:::             ||||||
    163 sLeuValProAlaGluSerMetThrAsn............IleGlnGlyL 176

    755 CCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGA 804
             ::::::|||::::::|||      :::|||||||||||||||
    176 euCysGlySerGlnHisAsnLys......SerAsnLeuThrMetGlu... 189

    805 GGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAA 854
           |||      |||      |||::::::
    190 ...AspValSerProGlyThrSerGlnMet..................... 198

    855 GAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATTTGG 904
         ||||||   :::   :::|||      ::::::      ||||||:::|
    199 .ArgAlaArgArgHisLysArgGluThrLeuLysMetThrLysTyrValG 215

    905 AACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCT 954

        ||||||:::||||||   |||::::::   ::::::::::::   ::::::
    215 luLeuValIleValAlaAspAsnArgGluPheGlnArgGlnGlyLysAsp 231

    955 CATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTC 1004
                ::::::::         :::::::::|||   ||||||::
    232 LeuGluLysValLysGlnArgLeuIleGluIleAlaAsnHisValAspLy 248

   1005 TATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGA 1054
        ::::|||:::      ||||||   |||:::||||||||||:::||||||:
    248 sPheTyrArg...ProLeuAsnIleArgIleValLeuValGlyValGluV 264

   1055 CCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAGATG 1104
        ::|||::::::   |||:::   :::|||:::   :::|||
    264 alTrpAsnAspIleAspLysCysSerIleSerGlnAspProPheThrArg 280

   1105 CTCCATGAGTTCTCAAAATACCGGCAGCGCATTAAG.....CA 1142
        ||||||||||||      :::|||   :::||||||           ::
    281 LeuHisGluPheLeuAspTrpArg...LysIleLysLeuLeuProArgLy 296

   1143 GCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGA 1192
        :      |||   ::::::|||||||||   |||   |||:::
    296 sSerHisAspAsnAlaGlnLeuIleSerGlyValTyrPheGlnGlyThrT 313

   1193 GAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCT......CGCACAAGA 1236
             :::::::::   :::   ::::::|||:::      ::::::
    313 hrIleGlyMetAlaProIleMetSerMetCysThrAlaGluGlnSerGly 329

   1237 GGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACAAGTATT 1286
        ||:|||   :::::::::          |||:::::::::|||   :::||
    330 GlyValValMetAspHisSerAspSerProLeuGlyAlaAlaValThrLe 346

   1287 ATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCA 1336
        |:::::::::|||:::::::|||:::|||::::::   :::      :::|
    346 uAlaHisGluLeuGlyHisAsnPheGlyMetAsnHisAspThrLeuGluA 363

   1337 GAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCATGGAG 1386
        ||      :::|||      :::   :::   |||||||||||||||:::
    363 rgGlyCysSerCysArgMetAlaAlaGluLysGlyGlyCysIleMetAsn 379

   1387 ...GAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTT 1433
           :::||||||            ||||||:::||||||
    380 ProSerThrGlyPheProPheProMetValPheSerSerCysSerArgLy 396

   1434 GGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACA 1483
         :::   :::      |||:::::|||   |||   ||||||||||||
    396 sAspLeuGluAlaSerLeuGluLysGlyMetGlyMetCysLeuPheAsnL 413

   1484 GGCCA......ACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATAC 1527
          |||         :::   |||         :::|||||||||||||||
    413 euProGluValLysGlnAlaPheGlyGlyArgLysCysGlyAsnGlyTyr 429

   1528 GTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTATGG 1577
        ||||||   ||||||||||||||||||||         ||||||   ::
    430 ValGluGluGlyGluGluCysAspCysGlyGluProGluGluCysThrAs 446

   1578 ATTATGCTGTAAG......AAATGTTCCCTCTCCAACGGGGCTCACTGCA 1621
        :   ||||||::         |||:::|||:::      |||   |||:
    446 nArgCysCysAsnAlaThrThrCysThrLeuLysProAspAlaValCysA 463

   1622 GCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGG 1671
        ::   |||   ||||||::::::      |||   ::::::|||   |||
    463 laHisGlyGlnCysCysGluAsp......CysGlnLeuLysProProGly 477

   1672 TATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTAC 1721
             ||||||   ::: .  |||:::|||||:::   |||:::||||
    478 ThrAlaCysArgGlySerSerAsnSerCysAspLeuProGluPheCysTh 494

   1722 TGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATG 1771
        ||||   :::   :::||||||   |||::::::   :::||||||:::
    494 rGlyThrAlaProHisCysProAlaAsnValTyrLeuHisAspGlyHisP 511

   1772 CATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGA 1821
          |||:::      :::|||   ||||||||||||   |||:::|||:::
```

```
     511 roCysGlnGlyValAspGlyTyrCysTyrAsnGlyIleCysGlnThrHis 527

    1822 GACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAA 1871
         ::::::||||||      :::||||||      |||      :::
     528 GluGlnGlnCysValThrLeuTrpGlyProGlyAlaLysProAlaProGl 544

    1872 GTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTCCG 1921
          :::|||:::|||::::::|||:::   |||          ||||||||||
     544 yIleCysPheGluArgValAsnSerAlaGlyAspProTyrGlyAsnCysG 561

    1922 GGAAGGATGGA...GACCGGTGGATTCAGTGCAGCAAACATGATGTGTTC 1968
         |||||||||:::   :::   :::   :::|||:::   :::|||:::
     561 lyLysAspSerLysSerAlaPheAlaLysCysGluLeuArgAspAlaLys 577

    1969 TGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCA 2018
         ||||||   :::   |||   :::   :::      |||   ||||||
     578 CysGlyLysIleGlnCysGlnGlyGlyAlaSerArgProValIleGly.. 593

    2019 ACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAA......GGCC 2062
         ::::::      |||   |||::::::   :::|||      ||||
     594 .ThrAsnAlaValSerIleGluThrAsnIleProGlnGlnGluGlyGlyA 610

    2063 GGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGAT...ACG 2109
         ||::::::   |||   |||:::||||||   |||   ||||||
     610 rgIleLeu...CysArgGlyThrHisValTyrLeuGlyAspAspMetPro 625

    2110 GATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTG 2159
         |||   |||   |||      ||||||   |||:::   :::   :::||
     626 AspProGlyLeuValLeuAlaGlyThrLysCysAlaGluGlyLysIleCy 642

    2160 TTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTC 2209
         ||||:::|||:::|||   :::|||:::::::::::::::   :::|||
     642 sLeuAsnArgArgCysGlnAsnIleSerValPheGlyValHisLysCysA 659

    2210 CACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAAT 2259
           ::::::               |||   |||:::||||||||||:::|||
     659 laMetGln.............CysHisGlyArgGlyValCysAsnAsn 670

    2260 GAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGCAT 2309
         :::   :::|||   |||:::      ||||||
     671 ArgLysAsnCysHisCysGluAlaHisTrpAla..... 681

    2310 CCGGGATCCAGTTAGGAACCTTCACCCCCCC..... 2340
                              ||||||
     682 .....ProProPheCysAspLysPheGlyP 690

     341 ..........AAGGATGAAGGACCC...AAGGGTCCTAGTGCCACCAAT 2376
                  |||:::||||||   :::      :::      :::
     690 heGlyGlySerThrAspSerGlyProIleArgGlnAlaAspAsnGlnGly 706

    2377 CTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCTATTGTCCT 2426
         |||   :::|||               |||||||||:::   :::   ||
     707 LeuThrValGly.............IleLeuValSerIleLeuCysLe 718

    2427 TGGGGGCACAGGCTGG 2442
         |   ::::::|||:::
     718 uLeuAlaAlaGlyPhe 723

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY41695

seq_documentation_block:
ID   AAY41695 standard; Protein; 735 AA.
XX
AC   AAY41695;
XX
DT   07-DEC-1999  (first entry)
XX
DE   Human PRO545 protein sequence.
XX
KW   Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW   probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW   secreted protein; transmembrane protein.
XX
OS   Homo sapiens.
XX
PN   WO9946281-A2.
XX
PD   16-SEP-1999.
XX
PF   08-MAR-1999;   99WO-US05028.
XX
PR   10-MAR-1998;   98US-0077450.
PR   11-MAR-1998;   98US-0077632.
PR   11-MAR-1998;   98US-0077641.
PR   11-MAR-1998;   98US-0077649.
PR   12-MAR-1998;   98US-0077791.
PR   13-MAR-1998;   98US-0078004.
PR   17-MAR-1998;   98US-0040220.
PR   20-MAR-1998;   98US-0078886.
PR   20-MAR-1998;   98US-0078910.
PR   20-MAR-1998;   98US-0078936.
PR   20-MAR-1998;   98US-0078939.
PR   25-MAR-1998;   98US-0079294.
PR   26-MAR-1998;   98US-0079656.
PR   27-MAR-1998;   98US-0079663.
PR   27-MAR-1998;   98US-0079664.
PR   27-MAR-1998;   98US-0079689.
PR   27-MAR-1998;   98US-0079728.
PR   27-MAR-1998;   98US-0079786.
PR   30-MAR-1998;   98US-0079920.
PR   30-MAR-1998;   98US-0079923.
PR   31-MAR-1998;   98US-0080105.
PR   31-MAR-1998;   98US-0080107.
PR   31-MAR-1998;   98US-0080165.
PR   31-MAR-1998;   98US-0080194.
PR   01-APR-1998;   98US-0080327.
PR   01-APR-1998;   98US-0080328.
PR   01-APR-1998;   98US-0080333.
PR   01-APR-1998;   98US-0080334.
PR   08-APR-1998;   98US-0081049.
PR   08-APR-1998;   98US-0081070.
PR   08-APR-1998;   98US-0081071.
PR   09-APR-1998;   98US-0081195.
PR   09-APR-1998;   98US-0081203.
PR   09-APR-1998;   98US-0081229.
PR   15-APR-1998;   98US-0081817.
PR   15-APR-1998;   98US-0081838.
PR   15-APR-1998;   98US-0081952.
PR   15-APR-1998;   98US-0081955.
PR   21-APR-1998;   98US-0082568.
PR   21-APR-1998;   98US-0082569.
PR   22-APR-1998;   98US-0082700.
PR   22-APR-1998;   98US-0082704.
PR   22-APR-1998;   98US-0082804.
PR   23-APR-1998;   98US-0082767.
PR   23-APR-1998;   98US-0082796.
PR   27-APR-1998;   98US-0083336.
PR   28-APR-1998;   98US-0083322.
PR   29-APR-1998;   98US-0083392.
PR   29-APR-1998;   98US-0083495.
PR   29-APR-1998;   98US-0083496.
PR   29-APR-1998;   98US-0083499.
PR   29-APR-1998;   98US-0083500.
PR   29-APR-1998;   98US-0083545.
PR   29-APR-1998;   98US-0083554.
PR   29-APR-1998;   98US-0083558.
PR   29-APR-1998;   98US-0083559.
PR   30-APR-1998;   98US-0083742.
PR   05-MAY-1998;   98US-0084366.
PR   06-MAY-1998;   98US-0084414.
PR   06-MAY-1998;   98US-0084441.
PR   07-MAY-1998;   98US-0084598.
PR   07-MAY-1998;   98US-0084600.
PR   07-MAY-1998;   98US-0084627.
PR   07-MAY-1998;   98US-0084637.
PR   07-MAY-1998;   98US-0084639.
```

```
PR   07-MAY-1998;   98US-0084640.
PR   07-MAY-1998;   98US-0084643.
PR   13-MAY-1998;   98US-0085323.
PR   13-MAY-1998;   98US-0085338.
PR   13-MAY-1998;   98US-0085339.
PR   15-MAY-1998;   98US-0085573.
PR   15-MAY-1998;   98US-0085579.
PR   15-MAY-1998;   98US-0085580.
PR   15-MAY-1998;   98US-0085582.
PR   15-MAY-1998;   98US-0085689.
PR   15-MAY-1998;   98US-0085697.
PR   15-MAY-1998;   98US-0085700.
PR   15-MAY-1998;   98US-0085704.
PR   18-MAY-1998;   98US-0086023.
PR   22-MAY-1998;   98US-0086392.
PR   22-MAY-1998;   98US-0086414.
PR   22-MAY-1998;   98US-0086430.
PR   22-MAY-1998;   98US-0086486.
PR   28-MAY-1998;   98US-0087098.
PR   28-MAY-1998;   98US-0087106.
PR   28-MAY-1998;   98US-0087208.
PR   30-JUL-1998;   98US-0094651.
PR   11-SEP-1998;   98US-0100038.

PA   (GETH ) GENENTECH INC.
XX
PI   Wood WI,  Goddard A,  Gurney A,  Yuan J,  Baker KP,  Chen J;
XX
DR   WPI; 1999-551358/46.
DR   N-PSDB; AAZ33953.
XX
PT   New secreted and transmembrane polypeptides and their polynucleotides,
PT   useful for treating blood coagulation disorders, cancers and cellular
PT   adhesion disorders -
XX
PS   Claim 12; Fig 30; 530pp; English.
XX
CC   The present invention describes secreted and transmembrane polypeptides
CC   and their polynucleotides. The nucleotide sequences are useful as
CC   sources of probes, primers, for chromosome mapping, and for generation
CC   of antisense sequences. They can also be used to create transgenic
CC   animals. The proteins can be used to treat a variety of diseases and
CC   disorders, depending on their function. Diseases that may be treated
CC   include blood coagulation disorders, cancers and cellular adhesion
CC   disorders. They may also be used to raise antibodies. AAZ33891 to
CC   AAZ34338, and AAY41685 to AAY41774 represent polynucleotide and
CC   polypeptide sequence given in the exemplification of the present
CC   invention.
XX
SQ   Sequence   735 AA;


[redacted]ment_scores:
          Quality:  965.50              Length:    749
            Ratio:  2.122                 Gaps:     23
 Percent Similarity:  60.748   Percent Identity: 31.242

alignment_block:
 US-09-634-252A-2 x AAY41695   ..

 Align seg 1/1  to: AAY41695  from: 1  to: 735

     138 GCTTCTCGTCCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCC 187
         |||:::||||||   |||::::::   ||||||   |||:::   |||
       2 AlaAlaArgProLeuProValSerProAlaArgAlaLeuLeuLeuAla.L 18

     188 GCGCCTGGGGGGCTGCTGCGCCCAGCGCTCCGCAT.....TGG 225
           |||   :::      ||||||         :::          |||
      18 euAlaGlyAlaLeuLeuAlaProCysGluAlaArgGlyValSerLeuTrp 34

     226 AATGAAACTGCAGAAAAAAATTTGGGAGTCCTGGCAGATGAAGACAATAC 275
         |||
      35 Asn..................................................... 35

     276 ATTGCAACAGAATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGA 325
             |||:::   ::::::::::::      :::::::::   :::
      36 .......GlnGlyArgAlaAspGluValValSerAlaSerValArgSerG 50

     326 AAGAAATCACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCG 375
           ::::::   :::|||   :::          :::   ||||||
      50 lyAspLeuTrpIleProValLys..............SerPheAspSer 61

     376 GAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACA 425
         ::::::      :::|||||:::   :::   ::::::::::|||
      62 LysAsnHisProGluValLeuAsnIleArgLeuGlnArgGluSerLys.. 77

     426 TAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCG 475

      77 ..................................................... 77

     476 GCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCT 525
           ::::::|||:::::::|||      |||:::||||||::::::|||
      78 .....GluLeuIleIleAsnLeuGluArgAsnGluGlyLeuIleAlaSer 92

     526 GATTATGTGGAGATTCACTACGAAAATGGGAAACCACAGTACTCTAAGGG 575
         ::::::::::|||   ||||||   :::          :::   |||   ::
      93 SerPheThrGluThrHisTyrLeuGlnAspGlyThrAspValSerLeuAl 109

     576 TGGAGAG.........CACTGTTACTACCATGGAAGCATCAGAGGCGTCA 616
         :   :::          ||||||||||||||||||   :::||||||   :
     109 aArgAsnTyrThrGlyHisCysTyrTyrHisGlyHisValArgGlyTyrS 126

     617 AAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATG... 663
         ::|||||   |||:::||||||||||||:::||||||:::|||:::
     126 erAspSerAlaValSerLeuSerThrCysSerGlyLeuArgGlyLeuIle 142

     664 ...TTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGT 710
            ||||||:::::::::      |||:::::||||||:::
     143 ValPheGluAsnGluSer......TyrValLeuGluProMet........ 154

     711 TCATGATGAGAAAAGCACAGGTCGA.............CCACATATAATCC 748
               ::::::::::|||:::|||             |||
     155 .......LysSerAlaThrAsnArgTyrLysLeuProAlaLysLysL 169

     749 AGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATG 798
           |||::::::   |||:::   :::::::::   ::::::   ::::::
     169 euLysSerValArgGlySerCysGlySerHisHisAsnThrProAsnLeu 185

     799 GAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAG 848
                 :::   :::|||      |||:::      |||   :::||
     186 AlaAlaLysAsnValProProProSerGln...ThrTrpAlaArgAr 201

     849 AAGGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAAT 898
         |:::||||||   :::                    ::::::      ||||
     201 gHisLysArgGluThr..................LeuLysAlaThrLysT 212

     899 ATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGC 948
         ||:::||||||:::||||||   |||::::::   ::::::::::::
     212 yrValGluLeuValIleValAlaAspAsnArgGluPheGlnArgGlnGly 228

     949 TCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGT 998
         ::::::         :::::::::          :::::::::|||   ||
     229 LysAspLeuGluLysValLysGlnArgLeuIleGluIleAlaAsnHisVa 245

     999 GGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTG 1048
         ||||::::::|||:::      ||||||   |||:::|||||||||:::|
     245 lAspLysPheTyrArg...ProLeuAsnIleArgIleValLeuValGlyV 261

    1049 TAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTG 1098
         |||||:::|||::::::   |||:::   :::::::::   :::|||
     261 alGluValTrpAsnAspMetAspLysCysSerValSerGlnAspProPhe 277

    1099 CAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGC.............AT 1136
               ||||||||||||      :::|||:::
     278 ThrSerLeuHisGluPheLeuAspTrpArgLysMetLysLeuLeuProAr 294
```

```
1137 TAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACT 1186
      |||:::|||   |||   ::::::|||:::|||   |||   |||:::
 294 gLysSerHis...AspAsnAlaGlnLeuValSerGlyValTyrPheG.nG 310

1187 ATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCT......CGC 1230
               :::::::::   :::   ::::::|||:::       :::
 310 lyThrThrIleGlyMetAlaProIleMetSerMetCysThrAlaAspGln 326

1231 ACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACA 1280
      :::   |||:::   :::::::::          |||:::::::::|||
 327 SerGlyGlyIleValMetAspHisSerAspAsnProLeuGlyAlaAlaVa 343

1281 AGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTT 1330
      :::|||:::::::::|||:::::|||:::|||::::::   :::
 343 lThrLeuAlaHisGluLeuGlyHisAsnPheGlyMetAsnHisAspThrL 360

1331 CTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATC 1380
       :::|||      :::|||:::   :::   :::   ||||||||||||
 360 euAspArgGlyCysSerCysGlnMetAlaValGluLysGlyGlyCysIle 376

1381 ATGGAG...GAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAG 1427
      |||:::   :::||||||                  ||||||:::||||
 377 MetAsnAlaSerThrGlyTyrProPheProMetValPheSerSerCysSe 393

1428 CATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTT 1477
      |      :::   :::      |||:::::|||   |||:::|||||||
 393 rArgLysAspLeuGluThrSerLeuGluLysGlyMetGlyValCysLeuP 410

1478 TCAACAGGCCA......ACAAAGCTATTTGAGCCCACGGAATGTGGAAAT 1521
      |||||   |||         :::   |||         :::|||||||||
 410 heAsnLeuProGluValArgGluSerPheGlyGlyGlnLysCysGlyAsn 426

1522 GGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATG 1571
       :::||||||  ||||||||||||||||||||          |||||
 427 ArgPheValGluGluGlyGluGluCysAspCysGlyGluProGluGluCy 443

1572 CTATGGATTATGCTGTAAG......AAATGTTCCCTCTCCAACGGGGTTC 1615
      |   :::   ||||||:::         |||:::|||:::      |||
 443 sMetAsnArgCysCysAsnAlaThrThrCysThrLeuLysProAspAlaV 460

1616 ACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCA 1665
       |||:::   |||   ||||||:::::      |||   :::::|||
 460 alCysAlaHisGlyLeuCysCysGluAsp......CysGlnLeuLysPro 474

1666 CGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATA 1715
       |||      ||||||||:::   |||:::||||||:::   |||::
 475 AlaGlyThrAlaCysArgAspSerSerAsnSerCysAspLeuProGluPh 491

1716 TTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACG 1765
      :|||||||||   |||   :::||||||   |||:::::   :::||||
 491 eCysThrGlyAlaSerProHisCysProAlaAsnValTyrLeuHisAspG 508

1766 GATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAG 1815
      ||:::::|||:::::   :::|||   ||||||||||||   ||:::
 508 lyHisSerCysGlnAspValAspGlyTyrCysTyrAsnGlyIleCysGln 524

1816 ACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTC 1865
      |||:::::::::||||||      :::||||||      |||      ::
 525 ThrHisGluGlnGlnCysValThrLeuTrpGlyProGlyAlaLysProAl 541

1866 TGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAA 1915
      :      :::|||:::|||:::::::|||:::   |||         ||||
 541 aProGlyIleCysPheGluArgValAsnSerAlaGlyAspProTyrGlyA 558

1916 ACTGCGGGAAG...GATGGAGACCGGTGGATTCAGTGCAGCAAACATGAT 1962
      ||||||||||   :::   :::   :::   :::|||:::   :::|||
 558 snCysGlyLysValSerLysSerSerPheAlaLysCysGluMetArgAsp 574

1963 GTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTAT 2012
      :::   ||||||   :::   |||   :::   :::      |||   ||
 575 AlaLysCysGlyLysIleGlnCysGlnGlyGlyAlaSerArgProValIl 591

2013 TGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCC 2062
```

```
       ||||           :::|||      :::         :::||||||
 591 eGlyThrAsnAlaValSerIleGluThrAsnIleProLeuGlnGlnGlyG 608

2063 GGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGAT...ACG 2109
           |||   |||   |||:::||||||   |||   ||||||
 608 lyArgIleLeuCysArgGlyThrHisValTyrLeuGlyAspAspMetPro 624

2110 GATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTG 2159
      |||   |||   |||      ||||||   |||:::   :::   :::||
 625 AspProGlyLeuValLeuAlaGlyThrLysCysAlaAspGlyLysIleCy 641

2160 TTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTC 2209
      ||||:::|||:::|||   :::|||:::::::::::::::   :::|||
 641 sLeuAsnArgGlnCysGlnAsnIleSerValPheGlyValHisGluCysA 658

2210 CACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAAT 2259
       ::::::               |||   |||:::|||||||||:::|||
 658 laMetGln..............CysHisGlyArgGlyValCysAsnAsn 669

2260 GAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGC 2304
      :::   :::|||   |||:::      ||||||         |||
 670 ArgLysAsnCysHisCysGluAlaHisTrpAlaProProPheCys 684
```

```
seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAB44251

seq_documentation_block:
ID   AAB44251 standard; Protein; 735 AA.
XX
AC   AAB44251;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human PRO545 (UNQ346) protein sequence SEQ ID NO:74.
XX
KW   Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;
KW   expressed sequence tag; detection; cancer.
XX
OS   Homo sapiens.
XX
PN   WO200053756-A2.
XX
PD   14-SEP-2000.
XX
PF   18-FEB-2000; 2000WO-US04341.
XX
PR   08-MAR-1999;   99WO-US05028.
PR   12-MAR-1999;   99US-0123957.
PR   29-MAR-1999;   99US-0126773.
PR   21-APR-1999;   99US-0130232.
PR   28-APR-1999;   99US-0131445.
PR   14-MAY-1999;   99US-0134287.
PR   23-JUN-1999;   99US-0141037.
PR   26-JUL-1999;   99US-0145698.
PR   29-OCT-1999;   99US-0162506.
PR   30-NOV-1999;   99WO-US28313.
PR   02-DEC-1999;   99WO-US28551.
PR   02-DEC-1999;   99WO-US28565.
PR   16-DEC-1999;   99WO-US30095.
PR   30-DEC-1999;   99WO-US31243.
PR   30-DEC-1999;   99WO-US31274.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00277.
PR   06-JAN-2000; 2000WO-US00376.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Botstein D,  Desnoyers L,  Eaton DL;
PI   Ferrara N,  Filvaroff E,  Fong S,  Gao W,  Gerber H,  Gerritsen ME;
PI   Goddard A,  Godowski PJ,  Grimaldi CJ,  Gurney AL,  Hillan KJ;
PI   Kljavin IJ,  Kuo SS,  Napier MA,  Pan J,  Paoni NF,  Roy MA;
PI   Shelton DL,  Stewart TA,  Tumas D,  Williams PM,  Wood WI;
XX
DR   WPI; 2000-611443/58.
```

```
DR   N-PSDB; AAC78476.
XX
PT   Novel PRO polypeptides and polynucleotides used in detection methods,
PT   to target bioactive molecules to specific cells, and to modulate
PT   cellular activities -
XX
PS   Claim 12; Fig 30; 636pp; English.
XX
CC   AAC78458 to AAC78599 represent polynucleotide and EST (expressed
CC   sequence tag) sequences which encode secreted or transmembrane PRO
CC   polypeptides. The PRO polynucleotides and polypeptides have cytostatic
CC   activity. The polynucleotides and polypeptides can be used for detecting
CC   the presence of PRO polypeptides in samples, for linking bioactive
CC   molecules to cells and for modulating biological activities of cells,
CC   using the polypeptides for specific targeting. The polypeptide targeting
CC   can be used to kill the target cells, e.g. for the treatment of cancers.
CC   The polypeptide pairs provide specific targeting of bioactive molecules
CC   to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
CC   the isolation of the PRO polynucleotide sequences.
XX
     Sequence   735 AA;


    ıment_scores:
           Quality:  965.50              Length:    749
             Ratio:  2.122                 Gaps:     23
 Percent Similarity:  60.748   Percent Identity: 31.242

alignment_block:
 US-09-634-252A-2 x AAB44251   ..

 Align seg 1/1  to: AAB44251  from: 1  to: 735

     138 GCTTCTCGTCCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCC 187
         |||:::||||||   |||::::::   ||||||   |||:::   |||
           2 AlaAlaArgProLeuProValSerProAlaArgAlaLeuLeuLeuAla.L 18

     188 GCGCCTGGGGGGCTGCTGCGCCCAGCGCTCCGCAT.....TGG 225
           |||   :::      ||||||         :::           |||
          18 euAlaGlyAlaLeuLeuAlaProCysGluAlaArgGlyValSerLeuTrp 34

     226 AATGAAACTGCAGAAAAAAATTTGGGAGTCCTGGCAGATGAAGACAATAC 275
         |||
          35 Asn..................................................... 35

     276 ATTGCAACAGAATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGA 325
                |||:::   ::::::::::::      :::::::::   :::
           36 .......GlnGlyArgAlaAspGluValValSerAlaSerValArgSerG 50

     326 AAGAAATCACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCG 375
           ::::::   :::|||   :::               :::   ||||||
           50 lyAspLeuTrpIleProValLys.............SerPheAspSer 61

     376 GAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACA 425
         ::::::      :::||||||:::   :::   ::::::::::::|||
           62 LysAsnHisProGluValLeuAsnIleArgLeuGlnArgGluSerLys.. 77

     426 TAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCG 475

           77 ..................................................... 77

     476 GCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCT 525
               ::::::|||::::::|||      |||:::||||||::::::|||
           78 .....GluLeuIleIleAsnLeuGluArgAsnGluGlyLeuIleAlaSer 92

     526 GATTATGTGGAGATTCACTACGAAAATGGGAAACCACAGTACTCTAAGGG 575
         :::::::::|||   ||||||   :::          :::   |||   ::
           93 SerPheThrGluThrHisTyrLeuGlnAspGlyThrAspValSerLeuAl 109

     576 TGGAGAG.........CACTGTTACTACCATGGAAGCATCAGAGGCGTCA 616
         :   :::          ||||||||||||||||   :::||||||   :
     109 aArgAsnTyrThrGlyHisCysTyrTyrHisGlyHisValArgGlyTyrS 126

     617 AAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATG... 663
         ::||||||   |||:::||||||||||||:::||||||:::|||:::
     126 erAspSerAlaValSerLeuSerThrCysSerGlyLeuArgGlyLeuIle 142

     664 ...TTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGT 710
            ||||||::::::::::      |||::::::||||||:::
     143 ValPheGluAsnGluSer......TyrValLeuGluProMet........ 154

     711 TCATGATGAGAAAAGCACAGGTCGA............CCACATATAATCC 748
                 :::::::::|||:::|||           |||
     155 .......LysSerAlaThrAsnArgTyrLysLeuProAlaLysLysL 169

     749 AGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATG 798
           |||::::::   |||:::   :::::::::   ::::::   ::::::
     169 euLysSerValArgGlySerCysGlySerHisHisAsnThrProAsnLeu 185

     799 GAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAG 848
                   :::   :::|||      |||:::      |||   :::||
     186 AlaAlaLysAsnValPheProProProSerGln...ThrTrpAlaArgAr 201

     849 AAGGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAAT 898
         |:::||||||   :::                    ::::::      ||||
     201 gHisLysArgGluThr.................LeuLysAlaThrLysT 212

     899 ATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGC 948
         ||:::||||||:::||||||   |||::::::   ::::::::::::
     212 yrValGluLeuValIleValAlaAspAsnArgGluPheGlnArgGlnGly 228

     949 TCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGT 998
         ::::::          :::::::::          :::::::::|||   ||
     229 LysAspLeuGluLysValLysGlnArgLeuIleGluIleAlaAsnHisVa 245

     999 GGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTG 1048
         ||||:::::::|||:::      ||||||   |||:::|||||||||:::|
     245 lAspLysPheTyrArg...ProLeuAsnIleArgIleValLeuValGlyV 261

    1049 TAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTG 1098
         |||||:::|||::::::   |||:::   :::::::::   :::|||
     261 alGluValTrpAsnAspMetAspLysCysSerValSerGlnAspProPhe 277

    1099 CAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGC.............AT 1136
               ||||||||||||      :::|||:::
     278 ThrSerLeuHisGluPheLeuAspTrpArgLysMetLysLeuLeuProAr 294

    1137 TAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACT 1186
          |||:::|||   |||   ::::::|||:::|||   |||   |||:::
     294 gLysSerHis...AspAsnAlaGlnLeuValSerGlyValTyrPheGlnG 310

    1187 ATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCT......CGC 1230
                   :::::::::   :::   ::::::|||:::      :::
     310 lyThrThrIleGlyMetAlaProIleMetSerMetCysThrAlaAspGln 326

    1231 ACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACA 1280
         :::   |||:::   :::::::::          |||:::::::::|||
     327 SerGlyGlyIleValMetAspHisSerAspAsnProLeuGlyAlaAlaVa 343

    1281 AGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTT 1330
          :::|||::::::::::|||::::::|||:::|||::::::   :::
     343 lThrLeuAlaHisGluLeuGlyHisAsnPheGlyMetAsnHisAspThrL 360

    1331 CTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATC 1380
         :::|||      :::|||:::   :::   :::   ||||||||||||
     360 euAspArgGlyCysSerCysGlnMetAlaValGluLysGlyGlyCysIle 376

    1381 ATGGAG...GAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAG 1427
         |||:::   :::||||||                  ||||||:::|||||
     377 MetAsnAlaSerThrGlyTyrProPheProMetValPheSerSerCysSe 393

    1428 CATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTT 1477
         |     :::   :::      |||::::::|||   |||:::|||||||
     393 rArgLysAspLeuGluThrSerLeuGluLysGlyMetGlyValCysLeuP 410

    1478 TCAACAGGCCA......ACAAAGCTATTTGAGCCCACGGAATGTGGAAAT 1521
```

```
     410 heAsnLeuProGluValArgGluSerPheGlyGlyGlnLysCysGlyAsn 426

     1522 GGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATG 1571
          :::||||||   ||||||||||||||||||||          |||||
      427 ArgPheValGluGluGlyGluGluCysAspCysGlyGluProGluGluCy 443

     1572 CTATGGATTATGCTGTAAG......AAATGTTCCCTCTCCAACGGGGCTC 1615
          |   :::   ||||||:::         |||:::|||:::       |||
      443 sMetAsnArgCysCysAsnAlaThrThrCysThrLeuLysProAspAlaV 460

     1616 ACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCA 1665
          |||:::   |||   ||||||::::::      |||   ::::::|||
      460 alCysAlaHisGlyLeuCysCysGluAsp......CysGlnLeuLysPro 474

     1666 CGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATA 1715
          |||      |||||||||:::   |||:::||||||:::   |||::
      475 AlaGlyThrAlaCysArgAspSerSerAsnSerCysAspLeuProGluPh 491

     1716 TTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACG 1765
          :|||||||||   |||   :::||||||   |||::::::   :::||||
      491 eCysThrGlyAlaSerProHisCysProAlaAsnValTyrLeuHisAspG 508

     1766 GATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAG 1815
          ||:::::|||:::::   :::|||   ||||||||||   |||:::
      508 lyHisSerCysGlnAspValAspGlyTyrCysTyrAsnGlyIleCysGln 524

     1816 ACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTC 1865
          |||:::::::::|||||      :::||||||      |||       ::
      525 ThrHisGluGlnGlnCysValThrLeuTrpGlyProGlyAlaLysProAl 541

     1866 TGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAA 1915
          :      :::|||:::|||:::::::|||:::   |||         ||||
      541 aProGlyIleCysPheGluArgValAsnSerAlaGlyAspProTyrGlyA 558

     1916 ACTGCGGGAAG...GATGGAGACCGGTGGATTCAGTGCAGCAAACATGAT 1962
          ||||||||||   :::   :::   :::   :::|||:::   :::|||
      558 snCysGlyLysValSerLysSerSerPheAlaLysCysGluMetArgAsp 574

     1963 GTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTAT 2012
          :::   ||||||   :::   |||   :::   :::       |||   ||
      575 AlaLysCysGlyLysIleGlnCysGlnGlyGlyAlaSerArgProValIl 591

     2013 TGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGACC 2062
          ||||  .        :::|||      :::         :::||||||
      591 eGlyThrAsnAlaValSerIleGluThrAsnIleProLeuGlnGlnGlyG 608

     063 GGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGAT...ACG 2109
          |||   |||   |||:::||||||   |||   ||||||
      608 lyArgIleLeuCysArgGlyThrHisValTyrLeuGlyAspAspMetPro 624

     2110 GATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTG 2159
          |||   |||   |||      ||||||   |||:::   :::   :::||
      625 AspProGlyLeuValLeuAlaGlyThrLysCysAlaAspGlyLysIleCy 641

     2160 TTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTC 2209
          ||||:::|||:::|||   :::|||:::::::::::::::   :::|||
      641 sLeuAsnArgGlnCysGlnAsnIleSerValPheGlyValHisGluCysA 658

     2210 CACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAAT 2259
          ::::::                |||   |||:::||||||||:::|||
      658 laMetGln..............CysHisGlyArgGlyValCysAsnAsn 669

     2260 GAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGC 2304
          :::   :::|||   |||:::      ||||||         |||    .
      670 ArgLysAsnCysHisCysGluAlaHisTrpAlaProProPheCys 684

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW25719

seq_documentation_block:
ID   AAW25719 standard; Protein; 686 AA.
XX
AC   AAW25719;
XX
DT   08-MAY-1998  (first entry)
XX
DE   Human alpha meltrin protein fragment 2.
XX
KW   Meltrin; myotube formation; muscle tissue; therapy; membrane protein;
KW   myoblasts; adhesion; fusion; medicinal; muscle disorder; human.
XX
OS   Homo sapiens.
XX
PN   WO9731109-A1.
XX
PD   28-AUG-1997.
XX
PF   17-OCT-1996;   96WO-JP03017.
XX
PR   23-FEB-1996;   96JP-0061756.
XX
PA   (MOCH ) MOCHIDA PHARM CO LTD.
XX
PI   Fujisawa A,  Mizushima C,  Ogawa N,  Shirakawa K,  Yamakawa T;
XX
DR   WPI; 1997-435161/40.
DR   N-PSDB; AAT86011.
XX
PT   Meltrin proteins involved with myotube formation in muscle - contain
PT   metallo:proteinase and disintegrin domains
XX
PS   Claim 5; Fig 15; 262pp; Japanese.
XX
CC   AAW25718-W25720 represent fragments of a new human membrane protein,
CC   alpha meltrin, which participates in the adhesion and fusion of
CC   myoblasts in the course of myotube formation in muscle tissue. Meltrins
CC   and their partial peptides can be used in the study of myotube formation,
CC   and in medicinal compositions for muscle disorders. Meltrins contain an
CC   intracellular domain, a membrane-penetrating domain, a fusion peptide
CC   domain, a disintegrin domain, a metalloproteinase domain, a precursor
CC   domain and a cysteine rich domain although these are not shown in the
CC   specification. Meltrin antagonists can be used as bone absorption
CC   inhibitors, e.g. for the treatment of osteoporosis and hypercalcaemia
CC   and for prevention of cancer cell infiltration.
XX
SQ   Sequence   686 AA;


alignment_scores:
            Quality:  951.00              Length:    667
              Ratio:  2.264                 Gaps:     19
 Percent Similarity:  62.969   Percent Identity: 32.834

alignment_block:
 US-09-634-252A-2 x AAW25719   ..

 Align seq 1/1  to: AAW25719  from: 1  to: 686

     370 GACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCA 419
         ||||||:::::      :::|||||::   :::   ::::::::::
      11 AspSerLysAsnHisProGluValLeuAsnIleArgLeuGlnArgGluSe 27

     420 AAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAG 469
         :|||
      27 rLys..................................................... 28

     470 CCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTG 519
                    :::::::|||:::::|||      |||:::|||||||:::
      29 ...........GluLeuIleIleAsnLeuGluArgAsnGluGlyLeuIle 41

     520 TCTTCTGATTATGTGGAGATTCACTACGAAAATGGGAAACCACAGTACTC 569
         :::|||:::::::::|||   ||||||   :::          :::   ||
      42 AlaSerSerPheThrGluThrHisTyrLeuGlnAspGlyThrAspValSe 58

     570 TAAGGGTGGAGAG.........CACTGTTACTACCATGGAAGCATCAGAG 610
         |   :::   :::          ||||||||||||||||||   :::||||
```

```
 58 rLeuAlaArgAsnTyrThrGlyHisCysTyrTyrHisGlyHisValArgG 75

611 GCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGC 660
    ||   :::|||||||   |||:::|||||||||||||:::|||||||:::|||
 75 lyTyrSerAspSerAlaValSerLeuSerThrCysSerGlyLeuArgGly 91

661 ATG......TTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGA 704
    :::      ||||||:::::::::      |||:::::|||||||:::
 92 LeuIleGlyPheGluAsnGluSer......TyrValLeuGluProMet.. 105

705 GCTGGTTCATGATGAGAAAAGCACAGGTCGA............CCACATA 742
                :::::::::|||:::|||           |||
106 ............LysSerAlaThrAsnArgTyrLysLeuPheProAlaL 118

743 TAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTC 792
          |||::::::   |||:::   :::::::::   ::::::
118 ysLysLeuLysSerValArgGlySerCysGlySerHisHisAsnThrPro 134

793 ACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTT 842
    ::::::         :::   :::|||      |||:::      |||
135 AsnLeuAlaAlaLysAsnValPheProProProSerGln...ThrTrpAl 150

843 GAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAA 892
     :::|||:::||||||   :::                  ::::::
150 aArgArgHisLysArgGluThr.................LeuLysAlaT 161

893 TGAAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAG 942
      ||||||:::||||||:::||||||   |||::::::   :::::::::
161 hrLysTyrValGluLeuValIleValAlaAspAsnArgGluPheGlnArg 177

943 CATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAA 992
    :::   ::::::         :::::::::         ::::::::||
178 GlnGlyLysAspLeuGluLysValLysGlnArgLeuIleGluIleAlaAs 194

993 CCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGG 1042
    |   ||||||:::::::|||:::      ||||||   |||:::|||||||
194 nHisValAspLysPheTyrArg...ProLeuAsnIleArgIleValLeuV 210

1043 TGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAAC 1092
     ||:::||||||:::|||::::::   |||:::   :::::::::   :::
210 alGlyValGluValTrpAsnAspMetAspLysCysSerValSerGlnAsp 226

1093 CCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGC........ 1134
     |||         ||||||||||||      :::|||:::
227 ProPheThrSerLeuHisGluPheLeuAspTrpArgLysMetLysLeuLe 243

1135 ....ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACAT 1180
         |||:::|||   |||   :::::::|||:::|||   |||   |
243 uProArgLysSerHis...AspAsnAlaGlnLeuValSerGlyValTyrP 259

1181 TTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCT... 1227
     ||:::           :::::::::   :::   ::::::|||:::
259 heGlnGlyThrThrIleGlyMetAlaProIleMetSerMetCysThrAla 275

228 ...CGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGT 1274
       ::::::   |||:::   :::::::::         |||:::::::::
276 AspGlnSerGlyGlyIleValMetAspHisSerAspAsnProLeuGlyAl 292

1275 GGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGG 1324
     :|||   :::|||:::::::::|||:::::::|||:::|||:::::   :
292 aAlaValThrLeuAlaHisGluLeuGlyHisAsnPheGlyMetAsnHisA 309

1325 AACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGC 1374
     ::      :::|||      :::|||:::   :::   :::   ||||||
309 spThrLeuAspArgGlyCysSerCysGlnMetAlaValGluLysGlyGly 325

1375 TGCATCATGGAG...GAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAA 1421
     |||||||||:::   :::|||||||                ||||||::
326 CysIleMetAsnAlaSerThrGlyTyrProPheProMetValPheSerSe 342

1422 GTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCT 1471
     :||||||      :::   :::      |||:::::|||   |||:::|
342 rCysSerArgLysAspLeuGluThrSerLeuGluLysGlyMetGlyValC 359

1472 GCCTTTTCAACAGGCCA......ACAAAGCTATTTGAGCCCACGGAATGT 1515
     ||||||||||   |||         :::   |||         :::|||
359 ysLeuPheAsnLeuProGluValArgGluSerPheGlyGlyGlnLysCys 375

1516 GGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGT 1565
     ||||||   :::||||||   ||||||||||||||||||||
376 GlyAsnArgPheValGluGluGlyGluGluCysAspCysGlyGluProGl 392

1566 GGAATGCTATGGATTATGCTGTAAG......AAATGTTCCCTCTCCAACG 1609
      ||||||  :::   ||||||:::         |||:::|||:::
392 uGluCysMetAsnArgCysCysAsnAlaThrThrCysThrLeuLysProA 409

1610 GGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTT 1659
       |||   |||:::   |||   ||||||::::::      |||   :::
409 spAlaValCysAlaHisGlyLeuCysCysGluAsp......CysGlnLeu 423

1660 CAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTAC 1709
     :::|||   |||      ||||||||||:::   |||:::||||||:::
424 LysProAlaGlyThrAlaCysArgAspSerSerAsnSerCysAspLeuPr 440

1710 TGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGC 1759
      |||:::|||||||||   |||   :::||||||   |||::::::   :
440 oGluPheCysThrGlyAlaSerProHisCysProAlaAsnValTyrLeuH 457

1760 AAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAG 1809
     ::||||||::::::|||::::::   :::|||   ||||||||||.
457 isAspGlyHisSerCysGlnAspValAspGlyTyrCysTyrAsnGlyIle 473

1810 TGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGC 1859
     |||:::|||:::::::::||||||      :::||||||      |||
474 CysGlnThrHisGluGlnGlnCysValThrLeuTrpGlyProGlyAlaLy 490

1860 AGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGA 1909
        :::      :::|||:::|||::::::|||:::   |||
490 sProAlaProGlyIleCysPheGluArgValAsnSerAlaGlyAspProT 507

1910 AGGGAAACTGCGGGAAG...GATGGAGACCGGTGGATTCAGTGCAGCAAA 1956
       ||||||||||||||   :::   :::   :::   :::|||:::
507 yrGlyAsnCysGlyLysValSerLysSerSerPheAlaLysCysGluMet 523

1957 CATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCC 2006
     :::|||:::   ||||||   :::   |||   :::   :::      ||
524 ArgAspAlaLysCysGlyLysIleGlnCysGlnGlyGlyAlaSerArgPr 540

2007 ACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATC 2056
     |   ||||||              :::|||      :::          :::|
540 oValIleGlyThrAsnAlaValSerIleGluThrAsnIleProLeuGlnG 557

2057 AAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGAT 2106
     |||||      |||   |||   |||:::||||||   |||   ||||||
557 lnGlyGlyArgIleLeuCysArgGlyThrHisValTyrLeuGlyAspAsp 573

2107 ...ACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTAT 2153
        |||   |||   |||      ||||||   |||:::   :::
574 MetProAspProGlyLeuValLeuAlaGlyThrLysCysAlaAspGlyLy 590

2154 GATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCA 2203
      :::||||||:::|||:::|||   :::|||:::::::::::::   :
590 sIleCysLeuAsnArgGlnCysGlnAsnIleSerValPheGlyValHisG 607

2204 GCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGT 2253
     ::|||   ::::::               |||   |||:::|||||||||
607 luCysAlaMetGln..............CysHisGlyArgGlyValCys 618

2254 AGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTG 2303
     :::|||:::   :::|||   |||:::      ||||||         ||
619 AsnAsnArgLysAsnCysHisCysGluAlaHisTrpAlaProProPheCy 635

2304 C 2304
     |
635 s 635
```

```
seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB47106

seq_documentation_block:
ID   AAB47106 standard; Protein; 812 AA.
XX
AC   AAB47106;
XX
DT   16-MAY-2001  (first entry)
XX
DE   Second splice variant of MAPP.
XX
KW   Splice variant; zdint2; mammalian adhesion protease peptide; MAPP;
KW   testis; ovary; prostate; small intestine; colon; stomach; thyroid;
KW   spinal cord; lymph node; trachea; heart; wound healing; apoptosis;
     neurogenesis; tumor proliferation; ischemia reperfusion; inflammation;
     immunologic recognition; gamete maturation; platelet aggregation;
     infarction; brain; cancer; Alzheimer's disease; multiple sclerosis;
     congestive heart failure.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          1
FT                   /note= "Start of secretory peptide domain"
FT   Domain          25..30
FT                   /note= "End of secretory peptide domain"
FT   Domain          28..31
FT                   /note= "Start of putative propeptide domain"
FT   Domain          200
FT                   /note= "Possible end of putative propeptide domain"
FT   Domain          203
FT                   /note= "Possible end of putative propeptide domain"
FT   Domain          205
FT                   /note= "Possible end of putative propeptide domain"
FT   Domain          207
FT                   /note= "Possible end of putative propeptide domain"
FT   Domain          204
FT                   /note= "Possible start of protease domain"
FT   Domain          208
FT                   /note= "Possible start of protease domain"
FT   Active-site     345..356
FT                   /note= "Active zinc catalytic site"
FT   Domain          410
FT                   /note= "Possible end of protease domain"
FT   Domain          419
FT                   /note= "Possible end of protease domain"
FT   Domain          419..420
FT                   /note= "Start of disintegrin domain"
FT   Domain          495..496
FT                   /note= "End of disintegrin domain"
FT   Domain          475..488
FT                   /note= "Disintegrin loop domain"
FT   Binding-site    481..483
FT                   /note= "Analogous to a RGD binding loop"
FT   Domain          496..497
FT                   /note= "Start of cysteine-rich domain"
FT   Peptide         497..701
FT                   /note= "Claimed peptide fragment"
FT   Domain          701
FT                   /note= "End of cysteine-rich domain"
FT   Domain          702.724
FT                   /note= "Transmembrane domain, claimed"
FT   Domain          725.812
FT                   /note= "Cytoplasmic domain, claimed"
XX
PN   WO200109293-A2.
XX
PD   08-FEB-2001.
XX
PF   02-AUG-2000; 2000WO-US21085.
XX
PR   03-AUG-1999;   99US-0368070.
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Sheppard PO,  Baindur N,  Bishop PD;
XX
DR   WPI; 2001-202662/20.
DR   N-PSDB; AAC85514.
XX
PT   Mammalian adhesion protease peptides useful for delivery of therapeutic
PT   agents, for identifying agonists and antagonists and treating disorders
PT   of brain, heart tissue and platelet aggregation -
XX
PS   Claim 2; Page 91-96; 106pp; English.
XX
CC   This sequence is encoded by a splice variant of the zdint2 cDNA and
CC   represents an isolated mammalian adhesion protease peptide (MAPP).
CC   Analysis of tissue distribution of MAPP showed a transcript of
CC   approx. 4.4kb with a strong signal in testes, ovary, prostate,
CC   small intestine and colon, and a fainter signal in stomach, thyroid,
CC   spinal cord, lymph node and trachea. Also there were two transcripts,
CC   approx. 4kb and approx. 4.4kb, both showing medium strength signal in
CC   heart tissue.  Peptide fragments of MAPP are useful for modulating
CC   cell-cell interactions, where the cells are derived from tissues
CC   showing a strong level of expression.  MAPP proteins are useful for
CC   delivery of therapeutic agents such as proteases, radionuclides,
CC   chemotherapy agents and small molecules and for identifying cells,
CC   tissues, cell lines which respond to a MAPP-stimulated pathway. MAPP
CC   polypeptides are also useful for identifying agonists and antagonists,
CC   to prepare antibodies and to study proliferation or differentiation in
CC   e.g. testes, ovary, prostate, small intestine, colon and spinal cord.
CC   The disintegrin loop of MAPP is useful in assays and treatment of
CC   disorders of the spinal cord, heart, aorta, colon, bladder, small
CC   intestine, uterus, prostate, stomach, testis, ovary, mammary gland,
CC   appendix, lung, trachea, fetal lung and placenta. MAPP polypeptides,
CC   antagonists and agonists are useful for promoting wound healing,
CC   studying cell-cell interactions, apoptosis, neurogenesis, tumor
CC   proliferation, repair and remodeling of ischemia reperfusion and
CC   inflammation in vitro and in vivo, also useful in repair and remodeling
CC   after an ischemic event, modulating immunologic recognition, gamete
CC   maturation and/or platelet aggregation. Agonists are useful in
CC   promoting the growth and development of cells of the myeloid and
CC   lymphoid lineages in culture. Antibodies to MAPP are useful for tagging
CC   cells that express MAPP, isolating MAPP by affinity purification,
CC   determining circulating levels of MAPP polypeptides, for detecting or
CC   quantitating soluble MAPP as marker of underlying pathology or disease
CC   and as neutralizing antibodies or as antagonists to block MAPP in vitro
CC   and in vivo. MAPP polypeptides, nucleic acids and/or antibodies are
CC   useful in the treatment of disorders associated with infarct in brain
CC   or heart tissue and/or platelet aggregation and as diagnostic
CC   indicators of cancer. Disorders diagnosed or treated or prevented with
CC   MAPP polypeptides, include Alzheimer's disease, tumor formation,
CC   multiple sclerosis, congestive heart failure, ischemic reperfusion or
CC   infarct and degenerative diseases.
XX
SQ   Sequence   812 AA;

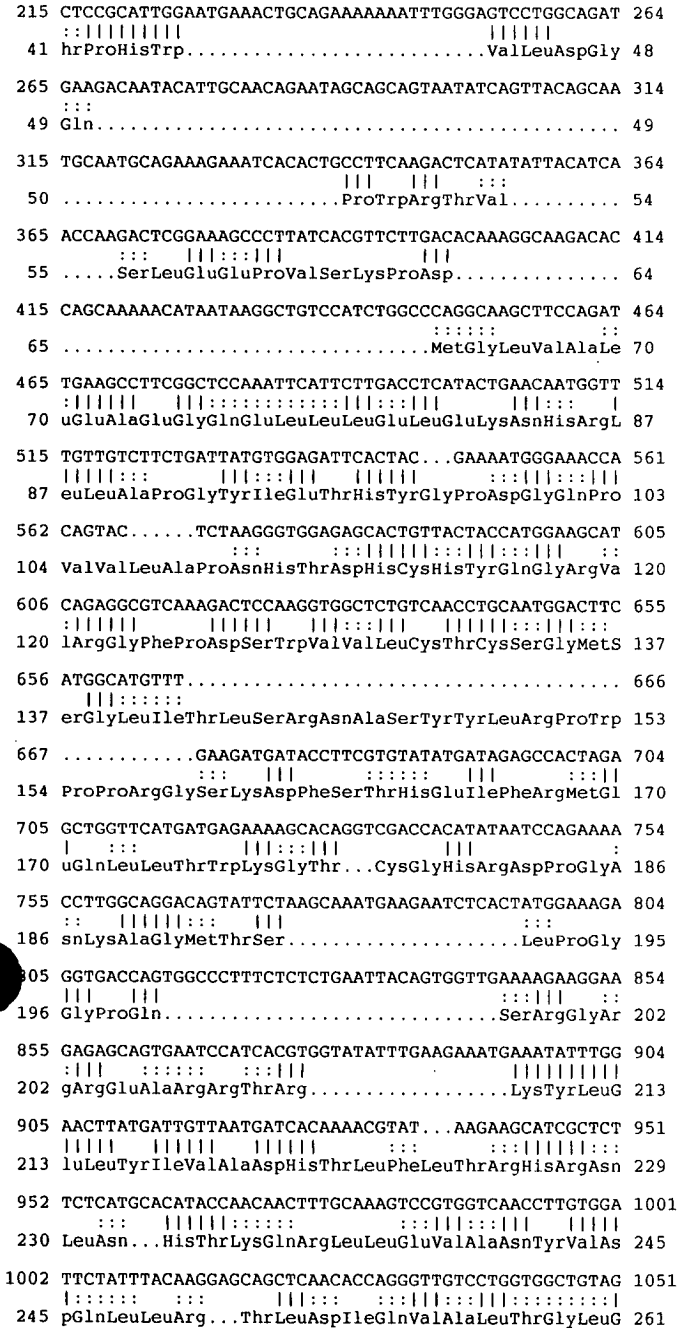

alignment_scores:
            Quality:  897.00              Length:    849
              Ratio:  1.971                 Gaps:     32
 Percent Similarity:  53.592   Percent Identity: 29.564

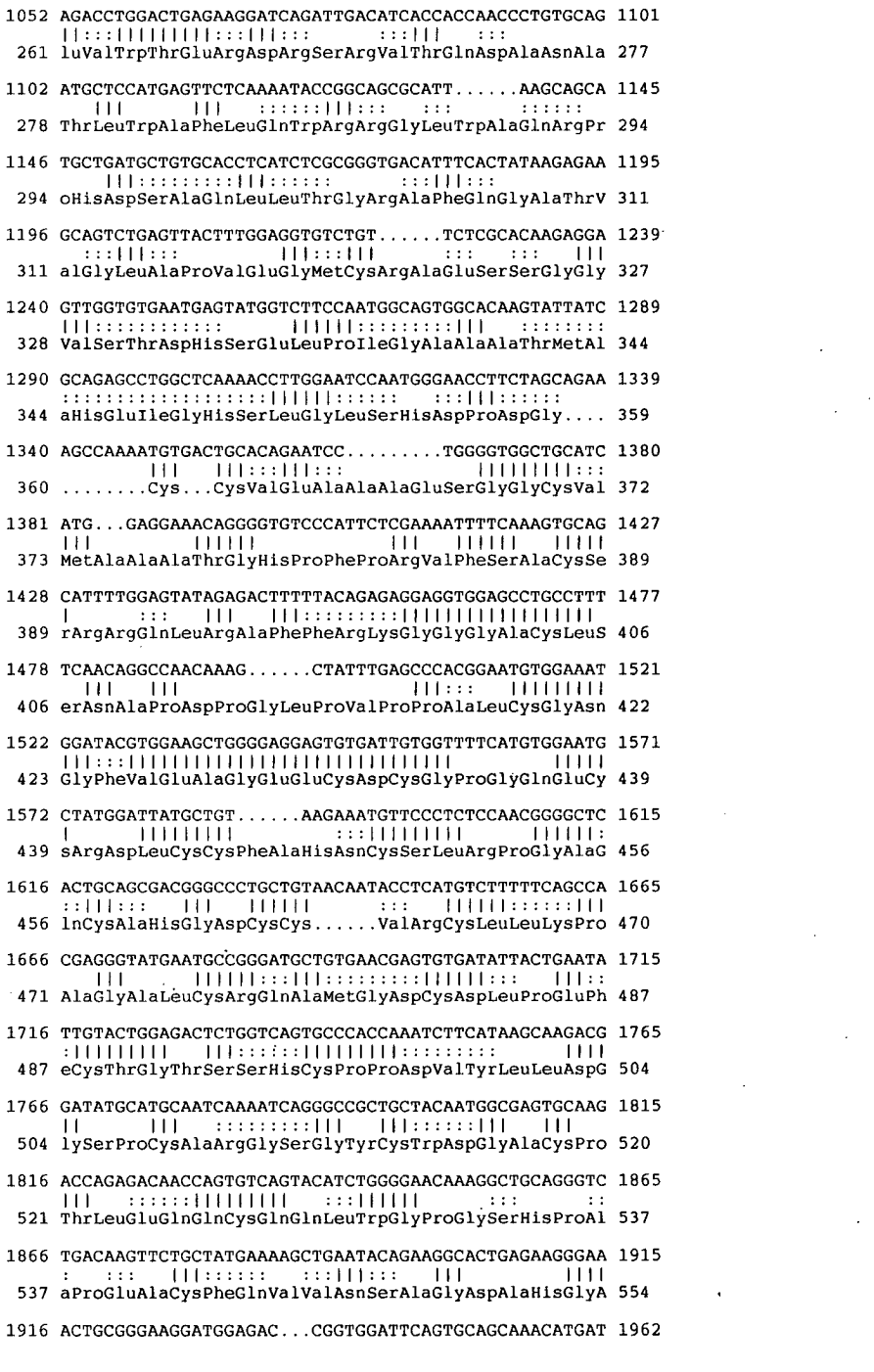
alignment_block:
 US-09-634-252A-2 x AAB47106   ..

 Align seg 1/1  to: AAB47106  from: 1  to: 812

     115 GCCCGCACGCCGCCCTGCCGCCTGCTTCTCGTCCTTCTCCTGCTGCCTCC 164
         ||||||      ||      |||||||||:::||||||||||||   ||
       8 AlaArgGlyThrProLeuLeuLeuLeuLeuLeuLeuLeuLeuTrpPr 24

     165 GCTCGCCGCCTCGTCCCGGCCCCGCGCCTGGGGGGCTGCTGCGCCCAGCG 214
         |:::   :::::::::      ::::::         :::   |||   :
      24 oValProGlyAlaGlyValLeuGlnGlyHisIleProGlyGlnProValT 41
```





```
     |||||||:::|||:::::::   :::::::::   |||:::   :::|||
  554 snCysGlyGlnAspSerGluGlyHisPheLeuProCysAlaGlyArgAsp 570

 1963 GTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTAT 2012
      ::::::||||||   |||   |||          ::::::   :::
  571 AlaLeuCysGlyLysLeuGlnCys...........GlnGlyGlyLysPr 583

 2013 TGGTCAACTTCAGGGTGAGATCATTCCA......ACTTCCTTCTACCATC 2056
       :::   |||      ::::::::::|||      ::::::   :::   :
  583 oSerLeuLeuAlaProHisMetValProValAspSerThrValHisLeuA 600

 2057 AAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGAT 2106
      ::|||:::   :::   |||   ||||||   ::::::   :::   :::
  600 spGlyGlnGluValThrCysArgGlyAlaLeuAlaLeuProSerAlaGln 616

  107 ACGGAT.........GTGGGCTATGTAGAAGATGGAACGCCATGTGGCCC 2147
         |||         :::|||   ||||||   ||||||   ||||||||
  617 LeuAspLeuLeuGlyLeuGlyLeuValGluProGlyThrGlnCysGlyPr 633

 2148 GTCTATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATA 2197
      |   |||:::|||   :::|||:::|||   :::              ::::
  633 oArgMetValCysGlnSerArgArgCysArgLysAsnAlaPheGlnGluL 650

 2198 TGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGG 2247
      :::::   |||                  :::|||   :::||| ||
  650 euGlnArgCysLeu..............ThrAlaCysHisSerHisGly 661

 2248 GTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGAC 2297
      ||||||:::::::::   :::|||   |||         ||||||
  662 ValCysAsnSerAsnHisAsnCysHisCysAlaProGlyTrpAlaProPr 678

 2298 AGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAAGGATG 2347
          |||:::                  ::::::            |||:
  678 oPheCysAsp............LysProGlyPheGlyGlySerMetAspS 691

 2348 AAGGACCCAAGGGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCT 2397
      ::||||||         :::|||   |||                :::|||
  691 erGlyPro......ValGlnAlaGluAsnHisAspThrPheLeuLeuAla 705

 2398 GGTGCCATCCTGGTAGCAGCTATTGTCCTTGGGGGCACAGGC........ 2439
           :::   |||        :::|||   |||:::|||
  706 MetLeuLeuSerValLeuLeuProLeuLeuProGlyAlaGlyLeuAlaTr 722

 2440 .....................................TGGGGAT 2446
                                          ||||||
  722 pCysCysTyrArgLeuProGlyAlaHisLeuGlnArgCysSerTrpGlyC 739

  447 TTAAAAATGTCAAGAAGAGAAGGTTCGATCCTACTCAGCAAGGCCCC 2493
                         |||   ||||||:::   :::||||||
  739 ys.................ArgArgAspProAlaCysSerGlyPro 748
```

```
seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW25717

seq_documentation_block:
ID   AAW25717 standard; Protein; 845 AA.
XX
AC   AAW25717;
XX
DT   08-MAY-1998  (first entry)
XX
DE   Mouse gamma meltrin protein.
XX
KW   Meltrin; myotube formation; muscle tissue; therapy; membrane protein;
KW   myoblast; adhesion; fusion; medicinal; murine; mouse.
XX
OS   Mus sp.
XX
PN   WO9731109-A1.
XX
PD   28-AUG-1997.
XX
PF   17-OCT-1996;   96WO-JP03017.
XX
PR   23-FEB-1996;   96JP-0061756.
XX
PA   (MOCH ) MOCHIDA PHARM CO LTD.
XX
PI   Fujisawa A,  Mizushima C,  Ogawa N,  Shirakawa K,  Yamakawa T;
XX
DR   WPI; 1997-435161/40.
DR   N-PSDB; AAT86009.
XX
PT   Meltrin proteins involved with myotube formation in muscle - contain
PT   metallo:proteinase and disintegrin domains
XX
PS   Claim 4; Fig 4; 262pp; Japanese.
XX
CC   This sequence represents a new membrane protein, gamma meltrin, which
CC   participates in the adhesion and fusion of myoblasts in the course of
CC   myotube formation in muscle tissue. Meltrins and their partial peptides
CC   can be used in the study of myotube formation, and in medicinal
CC   compositions for muscle disorders. Meltrins contain an intracellular
CC   domain, a membrane-penetrating domain, a fusion peptide domain, a
CC   disintegrin domain, a metalloproteinase domain, a precursor domain and a
CC   cysteine rich domain although these are not shown in the specification.
CC   Meltrin antagonists can be used as bone absorption inhibitors, e.g. for
CC   the treatment of osteoporosis and hypercalcaemia and for prevention of
CC   cancer cell infiltration.
XX
SQ   Sequence   845 AA;
```

```
alignment_scores:
            Quality:  892.00              Length:    803
              Ratio:  1.943                 Gaps:     26
 Percent Similarity:  57.161   Percent Identity: 29.639

alignment_block:
 US-09-634-252A-2 x AAW25717   ..

 Align seg 1/1  to: AAW25717  from: 1  to: 845

     103 GCCAGCGCCCCGGCCCGCACGCCGCCCTGCCGCCTGCTTCTCGTCCTTCT 152
         |||   :::|||      :::         |||   |||||||:::
       5 AlaLeuSerProLeuAlaSerLeuArgLeuArgTrpLeuLeuAlaCysGl 21

     153 CCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCGCGCCTGGGGGGCTG 202
          ||||||   |||:::      :::::||||||
      21 yLeuLeuGlyProValLeuGluAlaGlyArgPro..............  32

     203 CTGCGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAAAAAAATTTGGGA 252

      32 ..................................................  32

     253 GTCCTGGCAGATGAAGACAATACATTGCAACAGAATAGCAGCAGTAATAT 302
                 |||   ::::::|||::::::   :::|||
      33 .........AspLeuGluGlnThrValHisLeuSerSer...........  42

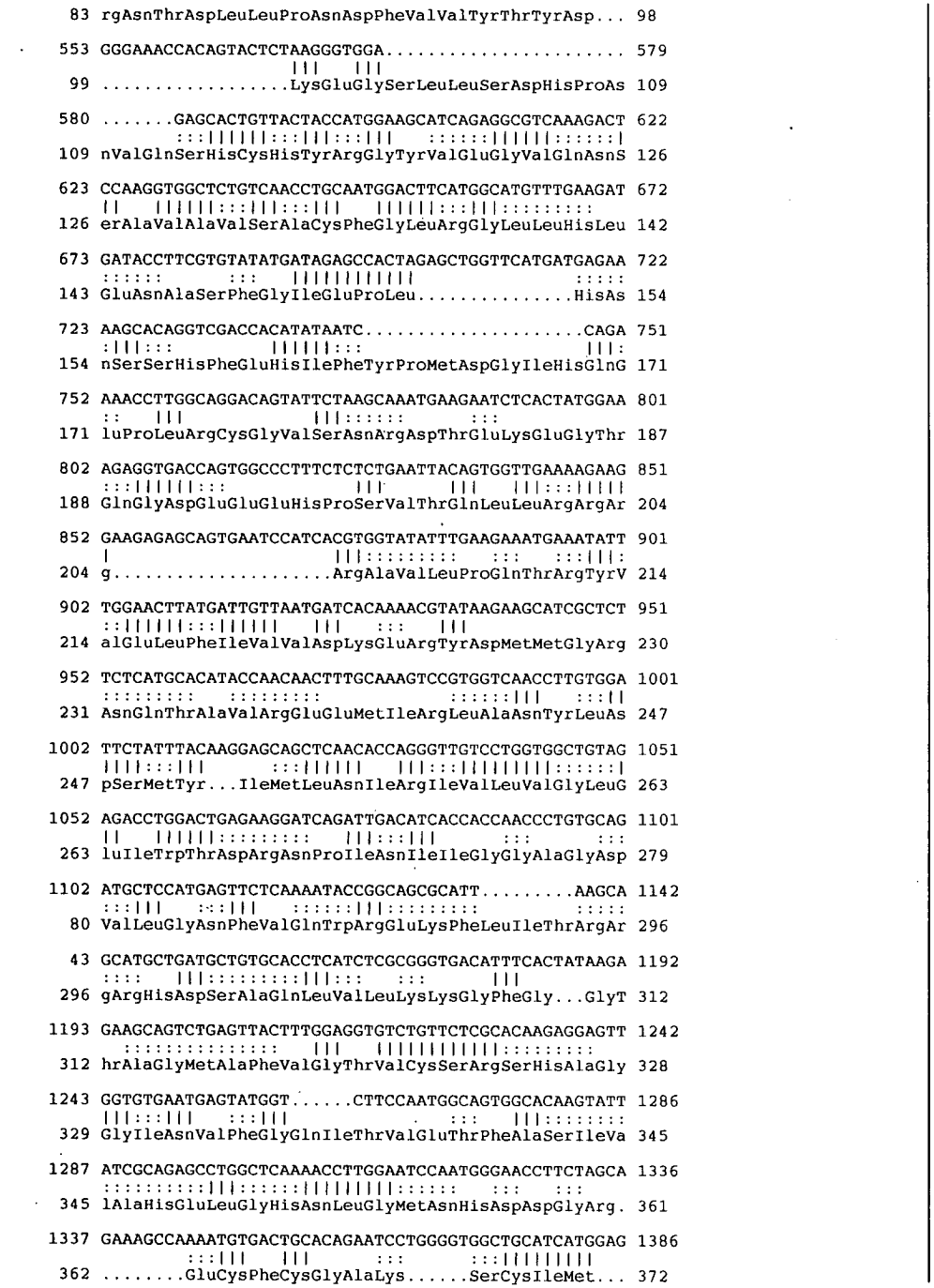
     303 CAGTTACAGCAATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCA 352
                                  ||||||      |||   ||||||
      43 .....................TyrGluIleIleThrProTrpArgLeu.  51

     353 TATATTACATCAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACA 402
                                                        |||
      52 ..................................................Thr  52

     403 AAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGC 452
         :::   |||::::::           ::::::           |||
      53 ArgGluArgArgGluAlaLeuGlyProSerSer.........GlnGlnIl  66

     453 AAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATAC 502
          |||:::   |||:::|||   |||::::::   |||:::   |||
      66 eSerTyrValIleGlnAlaGlnGlyLysGlnHisIleIleHisLeuGluA  83

     503 TGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAAAT 552
           |||:::   ||||||   :::|||:::|||         |||:::
```


83 rgAsnThrAspLeuLeuProAsnAspPheValValTyrThrTyrAsp... 98
553 GGGAAACCACAGTACTCTAAGGGTGGA..... 579
99 LysGluGlySerLeuLeuSerAspHisProAs 109
580 GAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACT 622
109 nValGlnSerHisCysHisTyrArgGlyTyrValGluGlyValGlnAsnS 126
623 CCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGAT 672
126 erAlaValAlaValSerAlaCysPheGlyLeuArgGlyLeuLeuHisLeu 142
673 GATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAA 722
143 GluAsnAlaSerPheGlyIleGluProLeu.....HisAs 154
723 AAGCACAGGTCGACCACATATAATC.....CAGA 751
154 nSerSerHisPheGluHisIlePheTyrProMetAspGlyIleHisGlnG 171
752 AAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAA 801
171 luProLeuArgCysGlyValSerAsnArgAspThrGluLysGluGlyThr 187
802 AGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAAG 851
188 GlnGlyAspGluGluGluHisProSerValThrGlnLeuLeuArgArgAr 204
852 GAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATT 901
204 g.....ArgAlaValLeuProGlnThrArgTyrV 214
902 TGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCT 951
214 alGluLeuPheIleValValAspLysGluArgTyrAspMetMetGlyArg 230
952 TCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGA 1001
231 AsnGlnThrAlaValArgGluGluMetIleArgLeuAlaAsnTyrLeuAs 247
1002 TTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAG 1051
247 pSerMetTyr...IleMetLeuAsnIleArgIleValLeuValGlyLeuG 263
1052 AGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAG 1101
263 luIleTrpThrAspArgAsnProIleAsnIleIleGlyGlyAlaGlyAsp 279
1102 ATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATT.....AAGCA 1142
80 ValLeuGlyAsnPheValGlnTrpArgGluLysPheLeuIleThrArgAr 296
43 GCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGA 1192
296 gArgHisAspSerAlaGlnLeuValLeuLysLysGlyPheGly...GlyT 312
1193 GAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTT 1242
312 hrAlaGlyMetAlaPheValGlyThrValCysSerArgSerHisAlaGly 328
1243 GGTGTGAATGAGTATGGT.....CTTCCAATGGCAGTGGCACAAGTATT 1286
329 GlyIleAsnValPheGlyGlnIleThrValGluThrPheAlaSerIleVa 345
1287 ATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCA 1336
345 lAlaHisGluLeuGlyHisAsnLeuGlyMetAsnHisAspAspGlyArg. 361
1337 GAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCATGGAG 1386
362 GluCysPheCysGlyAlaLys.....SerCysIleMet... 372
1387 GAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGA 1436
373 AsnSerGlyAlaSerGlySerArgAsnPheSerSerCysSerAlaGluAs 389
1437 GTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAAC.... 1482
389 pPheGluLysLeuThrLeuAsnLysGlyGlySerCysLeuLeuAsnIleP 406
1483 ..AGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATACGTG 1530
406 roLysProAspGluAlaTyrSerAlaProSerCysGlyAsnLysLeuVal 422
1531 GAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGC...TATGG 1577
423 AspProGlyGluGluCysAspCysGlyThrAlaLysGluCysGluValAs 439
1578 ATTATGCTGTAAG.....AAATGTTCCCTCTCCAACGGGGCTCACTGCA 1621
439 pProCysCysGluGlySerThrCysLysLeuLysSerPheAlaGluCysA 456
1622 GCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGG 1671
456 laTyrGlyAspCysCysLysAsp.....CysGlnPheLeuProGlyGly 470
1672 TATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTAC 1721
471 SerMetCysArgGlyLysThrSerGluCysAspValProGluTyrCysAs 487
1722 TGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATG 1771
487 nGlySerSerGlnPheCysProProAspValPheIleGlnAsnGlyTyrP 504
1772 CATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGA 1821
504 roCysGlnAsnSerLysAlaTyrCysTyrAsnGlyMetCysGlnTyrTyr 520
1822 GACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAA 1871
521 AspAlaGlnCysGlnValIlePheGlySerLysAlaLysAlaAlaProAr 537
1872 GTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCG 1921
537 gAspCysPheIleGluValAsnSerLysGlyAspArgPheGlyAsnCysG 554
1922 GGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGT 1971
554 lyPheSerGlySerGluTyrLysLysCysAlaThrGlyAsnAlaLeuCys 570
1972 GGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACT 2021
571 GlyLysLeuGlnCysGluAsnValGlnAspMetProValPheGly..... 585
2022 TCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTG 2071
586 IleValProAlaIleIleGlnThrProSerArgGlyThrL 599
2072 ACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACG...GATGTGGGC 2118
599 ysCysTrpGlyValAspPheGlnLeuGlySerAspValProAspProGly 615
2119 TATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCG 2168
616 MetValAsnGluGlyThrLysCysAspAlaGlyLysIleCysArgAsnPh 632
2169 GAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATT 2218
632 eGlnCysValAsnAlaSerValLeuAsnTyrAsp.....CysAspI 646
2219 CCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACC 2268
646 leGlnGlyLys...CysHisGlyHisGlyValCysAsnSerAsnLysAsn 661

```
    2269 TGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGA1CC 2318
         |||   |||:::      ||||||         |||:::
     662 CysHisCysGluAspGlyTrpAlaProProHisCysAsp..... 674

    2319 AGTTAGGAACCTTCACCCCCCCAAGGATGAAGGACCCAAG...GGTCCTA 2365
          :::::::::              |||:::||||||      :::   :
     675 .ThrLysGlyTyrGlyGlySerValAspSerGlyProThrTyrAsnAlaL 691

    2366 GTGCCACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCA 2415
         :::::|||   |||      |||   ::::::      :::|||::::::
     691 ysSerThrAlaLeuArgAspGlyLeuLeuValPhePhePheLeuIleVal 707

    2416 GCTATTGTC 2424
             :::|||
     708 ProLeuVal 710

     ame: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT:AAB47105

seq_documentation_block:
ID   AAB47105 standard; Protein; 802 AA.
XX
AC   AAB47105;
XX
DT   16-MAY-2001  (first entry)
XX
DE   First splice variant of MAPP.
XX
KW   Splice variant; zdint2; mammalian adhesion protease peptide; MAPP;
KW   testis; ovary; prostate; small intestine; colon; stomach; thyroid;
KW   spinal cord; lymph node; trachea; heart; wound healing; apoptosis;
KW   neurogenesis; tumor proliferation; ischemia reperfusion; inflammation;
KW   immunologic recognition; gamete maturation; platelet aggregation;
KW   infarction; brain; cancer; Alzheimer's disease; multiple sclerosis;
KW   congestive heart failure.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          1
FT                   /note= "Start of secretory peptide domain"
FT   Peptide         1..495
FT                   /note= "Claimed peptide fragment"
FT   Domain          25..30
FT                   /note= "End of secretory peptide domain"
FT   Domain          28..31
                     /note= "Start of putative propeptide domain"
     Peptide         31..495
                     /note= "Claimed peptide fragment"
     Peptide         31..200
                     /note= "Claimed peptide fragment"
FT   Domain          200
FT                   /note= "Possible end of putative propeptide domain"
FT   Domain          203
FT                   /note= "Possible end of putative propeptide domain"
FT   Domain          205
FT                   /note= "Possible end of putative propeptide domain"
FT   Domain          207
FT                   /note= "Possible end of putative propeptide domain"
FT   Domain          204
FT                   /note= "Possible start of protease domain"
FT   Domain          208
FT                   /note= "Possible start of protease domain"
FT   Peptide         208..410
FT                   /note= "Claimed peptide fragment"
FT   Peptide         208..495
FT                   /note= "Claimed peptide fragment"
FT   Active-site     345..356
FT                   /note= "Active zinc catalytic site"
FT   Domain          410
FT                   /note= "Possible end of protease domain"
FT   Domain          419
FT                   /note= "Possible end of protease domain"
FT   Domain          419..420
FT                   /note= "Start of disintegrin domain"
FT   Domain          495..496
FT                   /note= "End of disintegrin domain"
FT   Domain          475..488
FT                   /note= "Disintegrin loop domain, claimed"
FT   Binding-site    481..483
FT                   /note= "Analogous to a RGD binding loop"
FT   Domain          496..497
FT                   /note= "Start of cysteine-rich domain"
FT   Peptide         497..802
FT                   /note= "Claimed peptide fragment"
FT   Domain          802
FT                   /note= "End of cysteine-rich domain"
XX
PN   WO200109293-A2.
XX
PD   08-FEB-2001.
XX
PF   02-AUG-2000; 2000WO-US21085.
XX
PR   03-AUG-1999;   99US-0368070.
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Sheppard PO,  Baindur N,  Bishop PD;
XX
DR   WPI; 2001-202662/20.
DR   N-PSDB; AAC85513.
XX
PT   Mammalian adhesion protease peptides useful for delivery of therapeutic
PT   agents, for identifying agonists and antagonists and treating disorders
PT   of brain, heart tissue and platelet aggregation -
XX
PS   Claim 2; Page 83-88; 106pp; English.
XX
CC   This sequence is encoded by a splice variant of the zdint2 cDNA and
CC   represents an isolated mammalian adhesion protease peptide (MAPP).
CC   Analysis of tissue distribution of MAPP showed a transcript of
CC   approx. 4.4kb with a strong signal in testes, ovary, prostate,
CC   small intestine and colon, and a fainter signal in stomach, thyroid,
CC   spinal cord, lymph node and trachea. Also there were two transcripts,
CC   approx. 4kb and approx. 4.4kb, both showing medium strength signal in
CC   heart tissue.  Peptide fragments of MAPP are useful for modulating
CC   cell-cell interactions, where the cells are derived from tissues
CC   showing a strong level of expression.  MAPP proteins are useful for
CC   delivery of therapeutic agents such as proteases, radionuclides,
CC   chemotherapy agents and small molecules and for identifying cells,
CC   tissues, cell lines which respond to a MAPP-stimulated pathway. MAPP
CC   polypeptides are also useful for identifying agonists and antagonists,
CC   to prepare antibodies and to study proliferation or differentiation in
CC   e.g. testes, ovary, prostate, small intestine, colon and spinal cord.
CC   The disintegrin loop of MAPP is useful in assays and treatment of
CC   disorders of the spinal cord, heart, aorta, colon, bladder, small
CC   intestine, uterus, prostate, stomach, testis, ovary, mammary gland,
CC   appendix, lung, trachea, fetal lung and placenta. MAPP polypeptides,
CC   antagonists and agonists are useful for promoting wound healing,
CC   studying cell-cell interactions, apoptosis, neurogenesis, tumor
CC   proliferation, repair and remodeling of ischemia reperfusion and
CC   inflammation in vitro and in vivo, also useful in repair and remodeling
CC   after an ischemic event, modulating immunologic recognition, gamete
CC   maturation and/or platelet aggregation. Agonists are useful in
CC   promoting the growth and development of cells of the myeloid and
CC   lymphoid lineages in culture. Antibodies to MAPP are useful for tagging
CC   cells that express MAPP, isolating MAPP by affinity purification,
CC   determining circulating levels of MAPP polypeptides, for detecting or
CC   quantitating soluble MAPP as marker of underlying pathology or disease
CC   and as neutralizing antibodies or as antagonists to block MAPP in vitro
CC   and in vivo. MAPP polypeptides, nucleic acids and/or antibodies are
CC   useful in the treatment of disorders associated with infarct in brain
CC   or heart tissue and/or platelet aggregation and as diagnostic
CC   indicators of cancer. Disorders diagnosed or treated or prevented with
CC   MAPP polypeptides, include Alzheimer's disease, tumor formation,
CC   multiple sclerosis, congestive heart failure, ischemic reperfusion or
```

```
CC   infarct and degenerative diseases.
XX
SQ   Sequence   802 AA;


alignment_scores:
            Quality:  871.00              Length:    808
              Ratio:  2.007                Gaps:     30
 Percent Similarity:  53.713   Percent Identity: 28.960

alignment_block:
 US-09-634-252A-2 x AAB47105   ..

 Align seg 1/1  to: AAB47105  from: 1  to: 802

     115 GCCCGCACGCCGCCCTGCCGCCTGCTTCTCGTCCTTCTCCTGCTGCCTCC 164
         ||||||      |||      |||||||||:::|||||||||||||   ||
       8 AlaArgGlyThrProLeuLeuLeuLeuLeuLeuLeuLeuLeuTrpPr 24

     165 GCTCGCCGCCTCGTCCCGGCCCCGCGCCTGGGGGGCTGCTGCGCCCAGCG 214
         |:::   :::::::::      ::::::          :::   |||   :
      24 oValProGlyAlaGlyValLeuGlnGlyHisIleProGlyGlnProValT 41

     215 CTCCGCATTGGAATGAAACTGCAGAAAAAAATTTGGGAGTCCTGGCAGAT 264
         ::|||||||||                             ||||||
      41 hrProHisTrp.........................ValLeuAspGly 48

     265 GAAGACAATACATTGCAACAGAATAGCAGCAGTAATATCAGTTACAGCAA 314
         :::
      49 Gln.................................................. 49

     315 TGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACATCA 364
                                 |||   |||   :::
      50 ........................ProTrpArgThrVal.......... 54

     365 ACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACAC 414
               :::   |||:::|||             |||
      55 .....SerLeuGluGluProValSerLysProAsp.............. 64

     415 CAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGAT 464
                                         ::::::          ::
      65 ..............................MetGlyLeuValAlaLe 70

     465 TGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTT 514
         :||||||   |||:::::::::::|||:::|||      |||:::   |
      70 uGluAlaGluGlyGlnGluLeuLeuLeuGluLeuGluLysAsnHisArgL 87

     515 TGTTGTCTTCTGATTATGTGGAGATTCACTAC...GAAAATGGGAAACCA 561
         |||||:::      |||:::|||   ||||||      :::|||:::|||
      87 euLeuAlaProGlyTyrIleGluThrHisTyrGlyProAspGlyGlnPro 103

     562 CAGTAC......TCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCAT 605
                    :::      :::|||||||:::|||:::|||   ::
     104 ValValLeuAlaProAsnHisThrAspHisCysHisTyrGlnGlyArgVa 120

     606 CAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTC 655
         :||||||      ||||||   |||:::|||   |||||:::|||:::
     120 lArgGlyPheProAspSerTrpValValLeuCysThrCysSerGlyMetS 137

     656 ATGGCATGTTT.......................................... 666
           |||:::::::
     137 erGlyLeuIleThrLeuSerArgAsnAlaSerTyrTyrLeuArgProTrp 153

     667 ............GAAGATGATACCTTCGTGTATATGATAGAGCCACTAGA 704
                    :::   |||      ::::::   |||      :::||
     154 ProProArgGlySerLysAspPheSerThrHisGluIlePheArgMetGl 170

     705 GCTGGTTCATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAA 754
         |   :::         |||:::|||         |||            :
     170 uGlnLeuLeuThrTrpLysGlyThr...CysGlyHisArgAspProGlyA 186

     755 CCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGA 804
         ::   ||||||:::   |||                     :::

     186 snLysAlaGlyMetThrSer.....................LeuProGly 195

     805 GGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAA 854
         |||   |||                                 :::|||   ::
     196 GlyProGln.............................SerArgGlyAr 202

     855 GAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATTTGG 904
         :|||   ::::::   :::|||                   ||||||||||
     202 gArgGluAlaArgArgThrArg.................LysTyrLeuG 213

     905 AACTTATGATTGTTAATGATCACAAAACGTAT...AAGAAGCATCGCTCT 951
         |||||   ||||||   ||||||      :::      :::||||||:::
     213 luLeuTyrIleValAlaAspHisThrLeuPheLeuThrArgHisArgAsn 229

     952 TCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGA 1001
            :::   ||||||::::::         :::|||:::|||   |||||
     230 LeuAsn...HisThrLysGlnArgLeuLeuGluValAlaAsnTyrValAs 245

    1002 TTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAG 1051
         |:::::   :::      |||:::   :::|||:::|||:::::::::|
     245 pGlnLeuLeuArg...ThrLeuAspIleGlnValAlaLeuThrGlyLeuG 261

    1052 AGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAG 1101
         ||:::|||||||||:::|||:::      :::|||   :::
     261 luValTrpThrGluArgAspArgSerArgValThrGlnAspAlaAsnAla 277

    1102 ATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATT......AAGCAGCA 1145
            |||      |||   ::::::|||:::   :::      ::::::
     278 ThrLeuTrpAlaPheLeuGlnTrpArgArgGlyLeuTrpAlaGlnArgPr 294

    1146 TGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAA 1195
            |||:::::::::|||::::::      :::|||:::
     294 oHisAspSerAlaGlnLeuLeuThrGlyArgAlaPheGlnGlyAlaThrV 311

    1196 GCAGTCTGAGTTACTTTGGAGGTGTCTGT......TCTCGCACAAGAGGA 1239
         :::|||:::         |||:::|||      :::   :::   |||
     311 alGlyLeuAlaProValGluGlyMetCysArgAlaGluSerSerGlyGly 327

    1240 GTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACAAGTATTATC 1289
         |||:::::::::::::      ||||||:::::::::|||   ::::::::
     328 ValSerThrAspHisSerGluLeuProIleGlyAlaAlaAlaThrMetAl 344

    1290 GCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAA 1339
         ::::::::::::::::::::||||||::::::   :::|||::::::
     344 aHisGluIleGlyHisSerLeuGlyLeuSerHisAspProAspGly.... 359

    1340 AGCCAAAATGTGACTGCACAGAATCC.........TGGGGTGGCTGCATC 1380
                |||   |||:::|||:::            |||||||||:::
     360 ........Cys...CysValGluAlaAlaAlaGluSerGlyGlyCysVal 372

    1381 ATG...GAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAG 1427
         |||         ||||||           |||   ||||||   |||||
     373 MetAlaAlaAlaThrGlyHisProProArgValPheSerAlaCysSe 389

    1428 CATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTT 1477
         |      :::   |||   |||:::::::::|||||||||||||||||||
     389 rArgArgGlnLeuArgAlaPhePheArgLysGlyGlyGlyAlaCysLeuS 406

    1478 TCAACAGGCCAACAAAG......CTATTTGAGCCCACGGAATGTGGAAAT 1521
           |||   |||                     |||:::   ||||||||
     406 erAsnAlaProAspProGlyLeuProValProProAlaLeuCysGlyAsn 422

    1522 GGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATG 1571
         |||:::|||||||||||||||||||||||||||||      |||||
     423 GlyPheValGluAlaGlyGluGluCysAspCysGlyProGlyGlnGluCy 439

    1572 CTATGGATTATGCTGT......AAGAAATGTTCCCTCTCCAACGGGGCTC 1615
         |      |||||||||         :::|||||||||      ||||||:
     439 sArgAspLeuCysCysPheAlaHisAsnCysSerLeuArgProGlyAlaG 456

    1616 ACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCA 1665
         ::|||:::   |||   ||||||      :::   ||||||:::::::|||
     456 lnCysAlaHisGlyAspCysCys......ValArgCysLeuLeuLysPro 470
```

```
 1666 CGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATA 1715
       |||      ||||||:::|||:::::::::|||||:::   |||::
  471 AlaGlyAlaLeuCysArgGlnAlaMetGlyAspCysAspLeuProGluPh 487

 1716 TTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACG 1765
      :|||||||||   |||:::::::|||||||||:::::::::      ||||
  487 eCysThrGlyThrSerSerHisCysProProAspValTyrLeuLeuAspG 504

 1766 GATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAG 1815
      ||      |||   :::::::::|||   |||:::::::|||   |||
  504 lySerProCysAlaArgGlySerGlyTyrCysTrpAspGlyAlaCysPro 520

 1816 ACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTC 1865
      |||   ::::::|||||||||   :::||||||      :::      ::
  521 ThrLeuGluGlnGlnCysGlnGlnLeuTrpGlyProGlySerHisProAl 537

 1866 TGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAA 1915
      :   :::   |||:::::   :::|||:::   |||         ||||
  537 aProGluAlaCysPheGlnValValAsnSerAlaGlyAspAlaHisGlyA 554

 1916 ACTGCGGGAAGGATGGAGAC...CGGTGGATTCAGTGCAGCAAACATGAT 1962
      ||||||||:::|||:::::   :::::::::   |||:::   :::|||
  554 snCysGlyGlnAspSerGluGlyHisPheLeuProCysAlaGlyArgAsp 570

 1963 GTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGCAT 2012
      :::::|||||   |||   |||            ::::::   :::
  571 AlaLeuCysGlyLysLeuGlnCys.....GlnGlyGlyLysPr 583

 2013 TGGTCAACTTCAGGGTGAGATCATTCCA......ACTTCCTTCTACCATC 2056
       :::   |||      :::::::::|||      ::::::   :::   :
  583 oSerLeuLeuAlaProHisMetValProValAspSerThrValHisLeuA 600

 2057 AAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGAT 2106
      ::|||:::   :::   |||   ||||||   ::::::   :::   :::
  600 spGlyGlnGluValThrCysArgGlyAlaLeuAlaLeuProSerAlaGln 616

 2107 ACGGAT.........GTGGGCTATGTAGAAGATGGAACGCCATGTGGCCC 2147
         |||         :::|||   ||||||   ||||||   ||||||||
  617 LeuAspLeuLeuGlyLeuGlyLeuValGluProGlyThrGlnCysGlyPr 633

 2148 GTCTATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATA 2197
      |   |||:::|||   :::|||:::|||   :::            ::::
  633 oArgMetValCysGlnSerArgArgCysArgLysAsnAlaPheGlnGluL 650

 2198 TGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGG 2247
      :::::   |||                ::::||   :::||||||
  650 euGlnArgCysLeu..............ThrAlaCysHisSerHisGly 661

 2248 GTG..................................................... 2250
      :::
  662 AlaGlyLeuHisProSerValThrSerGlnAlaLeuValAlaAlaTrpTh 678

 2251 ..................................................TGTA 2254
                                                      ||||
  678 rValAlaLeuCysArgLeuLysThrMetThrProSerCysTrpProCysS 695

 2255 GTAATGAAGCCACCTGCATTTGTGATTTC......ACCTGGGCAGGGACA 2298
      ||:::   :::   |||:::|||:::         :::|||   |||:::
  695 erSerAlaSerCysCysLeuCysSerGlnGlyProAlaTrpProGlyVal 711

 2299 GATTGCAGTATCCGGGATCCAGTT 2322
            :::   ::::::|||:::
  712 AlaThrAspSerGlnGluProIle 719

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW01825

seq_documentation_block:
ID   AAW01825 standard; Protein; 751 AA.
XX
AC   AAW01825;
XX
DT   27-MAY-1997  (first entry)
XX
DE   Human metalloproteinase.
XX
KW   Metalloproteinase; infection; inflammation; allergy; therapy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          26..240
FT                   /label= Catalytic_domain
XX
PN   WO9709420-A2.
XX
PD   13-MAR-1997.
XX
PF   05-SEP-1996;   96WO-GB02180.
XX
PR   11-JUN-1996;   96GB-0012145.
PR   05-SEP-1995;   95GB-0018026.
PR   20-OCT-1995;   95GB-0021496.
PR   20-OCT-1995;   95GB-0021497.
PR   21-DEC-1995;   95GB-0026230.
XX
PA   (CLLT ) CELLTECH THERAPEUTICS LTD.
XX
PI   Docherty AJP,  Slocombe PM;
XX
DR   WPI; 1997-192896/17.
DR   N-PSDB; AAT62559.
XX
PT   DNA encoding human metallo:proteinase - used to develop prods. for
PT   treating, e.g. infectious diseases, inflammatory disease or allergy
XX
PS   Claim 5; Page 5; 17pp; English.
XX
CC   A novel human metalloproteinase (AAW01825), or its catalytic domain,
CC   can be expressed in prokaryotic or eukaryotic host cells using
CC   an isolated gene sequence (see also AAT62558-59).  The
CC   metalloproteinase can be used to screen for cpds. that regulate
CC   its activity, and to raise antibodies.  Such cpds. and antibodies
CC   are useful for the in vivo treatment of diseases or disorders
CC   related to over- or under-regulated prodn. of the enzyme.
XX
SQ   Sequence   751 AA;


alignment_scores:
            Quality:  829.00              Length:    548
              Ratio:  2.375                 Gaps:     15
 Percent Similarity:  63.686   Percent Identity: 33.394

alignment_block:
 US-09-634-252A-2 x AAW01825   ..

 Align seg 1/1  to: AAW01825  from: 1  to: 751

  835 CAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATT 884
      |||       |||:::|||   |||   :::       :::|||      ::
   17 GlnGlnThrLysLysArgProArgArgMet......LysArgGluAspLe 31

  885 TGAAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAACGT 934
      :::::::||||||||||:::||||||   :::|||   |||:::      :
   31 uAsnSerMetLysTyrValGluLeuTyrLeuValAlaAspTyrLeuGluP 48

  935 ATAAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCC 984
      :::::|||:::|||   ::::::      |||::::::          :::
   48 heGlnLysAsnArgArgAspGlnAspAlaThrLysHisLysLeuIleGlu 64

  985 GTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGT 1034
      :::::::|||   ||||||:::::::|||:::   :::|||||||   |||::
   65 IleAlaAsnTyrValAspLysPheTyrArg...SerLeuAsnIleArgIl 80

 1035 TGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCA 1084
```

```
     :::::||||||:::::::|||:::||||||:::   ::::::   :::::::
  80 eAlaLeuValGlyLeuGluValTrpThrHisGlyAsnMetCysGluValS 97

1085 CCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGC 1134
     ::   ||||||   :::   |||   :::|||   :::::::|||::::::
  97 erGluAsnProTyrSerThrLeuTrpSerPheLeuSerTrpArgArgLys 113

1135 ATT......AAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGAC 1178
     :::       :::::::::   |||   ::::::||||||:::   :::::
  114 LeuLeuAlaGlnLysTyrHisAspAsnAlaGlnLeuIleThrGlyMetSe 130

1179 ATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCT. 1227
     :||||||            :::|||:::   :::   ::::::||||||
  130 rPheHisGlyThrThrIleGlyLeuAlaProLeuMetAlaMetCysSerV 147

1228 .....CGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCA 1272
          ::::::   ||||||:::::::::::            ::::::
  147 alTyrGlnSerGlyGlyValAsnMetAspHisSerGluAsnAlaIleGly 163

1273 GTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATG 1322
     ||||||   :::::::::::::::::::::::::|||:::|||:::
  164 ValAlaAlaThrMetAlaHisGluMetGlyHisAsnPheGlyMetThrHi 180

1323 GGAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTG 1372
       :::   ::::::           |||:::::::::   :::   ||||
  180 sAspSerAlaAsp.........CysCysSerAlaSerAlaAlaAspGlyG 194

1373 GCTGCATCATG...GAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCA 1419
     ||||||||||         ||||||            :::   |||:::
  194 lyCysIleMetAlaAlaAlaThrGlyHisProPheProLysValPheAsn 210

1420 AAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGC 1469
        |||:::      |||         :::||||||   |||||||||
  211 GlyCysAsnArgArgGluLeuAspArgTyrLeuGlnSerGlyGlyGlyMe 227

1470 CTGCCTTTTCAACAGGCCA......ACAAAGCTATTTGAGCCCACGGAAT 1513
       ||||||   |||   |||          |||:::          :::|
  227 tCysLeuSerAsnMetProAspThrArgMetLeuTyrGlyGlyArgArgC 244

1514 GTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCAT 1563
     ||||||||||||||:::|||   ||||||||||||||||||||||   :::
  244 ysGlyAsnGlyTyrLeuGluAspGlyGluGluCysAspCysGlyGluGlu 260

1564 GTGGAATGCTATGGATTATGCTGT......AAGAAATGTTCCCTCTCCAA 1607
        ||||||   :::   ||||||      :::::::|||:::|||
  261 GluGluCysAsnAsnProCysCysAsnAlaSerAsnCysThrLeuArgPr 277

1608 CGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTT 1657
       ||||||:::|||:::   |||   |||||||:::       :::|||   :
  277 oGlyAlaGluCysAlaHisGlySerCysCysHis......GlnCysLysL 292

1658 TTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATT 1707
     ::         |||      |||||||:::   :::::::::|||||||:::
  292 euLeuAlaProGlyThrLeuCysArgGluGlnAlaArgGlnCysAspLeu 308

1708 ACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAA 1757
        |||:::||||||||   |||   :::||||||   |||:::::::::
  309 ProGluPheCysThrGlyLysSerProHisCysProThrAsnPheTyrGl 325

1758 GCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCG 1807
     ::::|||||      |||:::   :::|||:::   ||||||||||||
  325 nMetAspGlyThrProCysGluGlyGlyGlnAlaTyrCysTyrAsnGlyM 342

1808 AGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCT 1857
       |||   |||   ::::::|||||||||   :::||||||      |||
  342 etCysLeuThrTyrGlnGluGlnCysGlnGlnLeuTrpGlyProGlyAla 358

1858 GCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGA 1907
           :::      :::|||:::|||||||:::|||:::   |||
  359 ArgProAlaProAspLeuCysPheGluLysValAsnValAlaGlyAspTh 375

1908 GAAGGGAAACTGCGGGAAGGAT...GGAGACCGGTGGATTCAGTGCAGCA 1954
          ||||||||||||||||||   :::   :::      :::|||:::
  375 rPheGlyAsnCysGlyLysAspMetAsnGlyGluHisArgLysCysAsnM 392

1955 AACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCT 2004
        :::|||:::   ||||||   :::   |||   :::
  392 etArgAspAlaLysCysGlyLysIleGlnCysGlnSerSerGluAlaArg 408

2005 CCACGTATTGGTCAACTTCAGGGTGAGATCATTCCA......ACTTCCTT 2048
     |||            |||:::::::::   :::|||      |||:::::
  409 Pro...........LeuGluSerAsnAlaValProIleAspThrThrIl 421

2049 CTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTA. 2097
     :      :::||||||   |||:::|||   |||:::||||||
  421 eIleMetAsnGlyArgGlnIleGlnCysArgGlyThrHisValTyrArgG 438

2098 .:............GATGATGATACGGATGTGGGCTATGTAGAAGATGGA 2133
                    |||      |||   |||   |||      |||
  438 lyProGluGluGluGlyAspMetLeuAspProGlyLeuValMetThrGly 454

2134 ACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAAT 2183
     |||   ||||||   :::   :::|||::::::   :::|||   :::
  455 ThrLysCysGlyTyrAsnHisIleCysPheGluGlyGlnCysArgAsnTh 471

2184 TCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCT 2233
       :::   ::::::   ::::::|||                ||||||   |
  471 rSerPhePheGluThrGluGlyCys.............GlyLysLysC 483

2234 GTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTC 2283
     ||:::|||||||||||||||||:::|||:::   :::|||   |||
  483 ysAsnGlyHisGlyValCysAsnAsnAsnGlnAsnCysHisCysLeuPro 499

2284 ACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCA 2333
        ||||||         |||:::                  ::::::
  500 GlyTrpAlaProProPheCysAsnThrProGlyHisGlyGlySerIle.. 515

2334 CCCCCCCAAGGATGAAGGACCCAAGGGTCCTAGTGCCACCAATCTCATAA 2383
                |||:::||||||      |||:::::::::::   ::::
  516 ..........AspSerGlyProMetProProGluSerValGlyProValV 529

2384 TAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCTATTGTCCTT 2427
     ::::::::::      ||||||||||||   |||::::::::
  529 alAlaGlyValLeuValAlaIleLeuValLeuAlaValLeuMet 543
```

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAW90855

seq_documentation_block:

```
ID   AAW90855 standard; Protein; 775 AA.
XX
AC   AAW90855;
XX
DT   28-JUN-2000  (first entry)
XX
DE   Human ADAM protein #2.
XX
KW   ADAM protein; human; A disintegrin and metalloprotease; diagnosis;
KW   drug development; intervertebral hernia; sciatica; pulmonary fibrosis;
KW   diabetic nephropathy; hepatic fibrosis; glomerulitis; osteopetrosis.
XX
OS   Homo sapiens.
XX
PN   WO200014227-A1.
XX
PD   16-MAR-2000.
XX
PF   02-SEP-1999;   99WO-JP04766.
XX
PR   03-SEP-1998;   98JP-0250115.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Yoshimura K,  Hikichi Y,  Nishimura A;
XX
DR   WPI; 2000-271056/23.
DR   N-PSDB; AAZ89683.
```

```
XX
PT   Novel protein belong to A disintegrin and metalloprotease family, with
PT   protease activity and extracellular matrix digesting enzyme activity,
PT   for gene diagnosis and developing drugs for treating e.g. sciatica and
PT   glomerulitis -
XX
PS   Claim 4; Page 98-102; 109pp; Japanese.
XX
CC   This invention describes a novel human protein (I) which belongs to the
CC   ADAM (A disintegrin and metalloprotease) protein family. The protein,
CC   peptide fragment and antibody are useful for gene diagnosis and in the
CC   development of drugs to prevent or treat intervertebral hernia, sciatica,
CC   glomerulitis, diabetic nephropathy, hepatic fibrosis, pulmonary fibrosis
CC   or osteopetrosis. This sequence represents a human ADAM protein which is
CC   described in the method of the invention.

     Sequence   775 AA;


alignment_scores:
            Quality:  812.50              Length:    682
              Ratio:  2.031                Gaps:     26
 Percent Similarity:  58.651   Percent Identity: 31.232

alignment_block:
 US-09-634-252A-2 x AAW90855   ..

 Align seg 1/1  to: AAW90855  from: 1  to: 775

     328 GAAATCACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGA 377
         |||::::::   |||   ||||||
      29 GluValValTyrProIleArgLeu..........................  36

     378 AAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATA 427
             |||   |||      :::::::|||      :::||||||::::::
      37 .HisProLeuHisLysArgGluAlaLysGluProGluGlnGlnGluGlnP  53

     428 ATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGC 477
           :::            ::::::   ::::::::::::         |||
      53 heGlu............ThrGluLeuLysTyrLysMetThrIleAsnGly  65

     478 TCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGA 527
         :::      :::|||   |||      |||::::::||||||:::
      66 LysIleAlaValLeuTyrLeuLysLysAsnLysAsnLeuLeuAlaProGl  82

     528 TTATGTGGAGATTCACTACGAAAAT...GGGAAA..............C 559
          |||:::|||.  :::|||::::::   ||||||               |
      82 yTyrThrGluThrTyrTyrAsnSerThrGlyLysGluIleThrThrSerP  99

     560 CACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGA 609
         |||||               :::   |||||||||:::|||   |||
      99 roGlnIleMet.........AspAspCysTyrTyrGlnGlyHisIleLeu 112

     610 GGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGG 659
         :::   |||   |||   :::::::::|||||||||:::||||||:::||
     113 AsnGluLysValSerAspAlaSerIleSerThrCysArgGlyLeuArgGl 129

     660 CATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGG 709
         |   |||::::::           |||:::|||||||||||:::   :
     129 yTyrPheSerGlnGlyAspGlnArgTyrPheIleGluProLeuSerProI 146

     710 TTCATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAACCTTG 759
         ::|||   :::         |||:::   |||   :::   |||
     146 leHisArgAsp.........GlyGlnGluHisAlaLeuPheLys...... 157

     760 GCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAGGT.. 807
                 |||:::   :::   ||||||               |||
     158 .........TyrAsnProAspGluLysAsnTyrAspSerThrCysGlyMe 171

     808 .GACCAGTGGCCCTTTCTCTCTGAATTACAGTGG.........TTGAAAA 847
          |||         :::      :::||||||             |||
     171 tAspGlyValLeuTrpAlaHisAspLeuGlnGlnAsnIleAlaLeuProA 188

     848 GAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAA 897
             |||   :::          :::|||   :::   |||:::   |||
     188 laThrLysLeuValLysLeuLysAspArgLysValGlnGluHisGluLys 204

     898 TATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCG 947
         |||:::|||      :::|||   |||:::      :::|||::::::::
     205 TyrIleGluTyrTyrLeuValLeuAspAsnGlyGluPheLysArgTyrAs 221

     948 CTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTG 997
         ::::::::::   :::   ::::::   :::   ::::::::::|||   |
     221 nGluAsnGlnAspGluIleArgLysArgValPheGluMetAlaAsnTyrV 238

     998 TGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCT 1047
         ||:::   :::||||||   :::||||||||:::|||:::||||||:::
     238 alAsnMetLeuTyrLys...LysLeuAsnThrHisValAlaLeuValGly 253

    1048 GTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGT 1097
         :::|||   ||||||:::|||||||:::|||   ||||||   |||
     254 MetGluIleTrpThrAspLysAspLysIleLysIleThrProAsnAlaSe 270

    1098 GCAGATGCTCCATGAGTTCTCAAAATACCGG.............CAGCGCA 1135
                ||||::::::|||||||||:::|||              :::|||
     270 rPheThrLeuGluAsnPheSerLysTrpArgGlySerValLeuSerArgA 287

    1136 TTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCAC 1185
           |||:::|||   |||   ::::::||||||:::   :::   :::
     287 rgLysArgHis...AspIleAlaGlnLeuIleThrAlaThrGluLeuAla 302

    1186 TATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCT...CGCAC 1232
                    :::|||:::::::::::::   :::||||||      ::
     303 GlyThrThrValGlyLeuAlaPheMetSerThrMetCysSerProTyrSe 319

    1233 AAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACAAG 1282
         :   ||||||      ::::::            :::   ||||||   :
     319 rValGlyValValGlnAspHisSerAspAsnLeuLeuArgValAlaGlyT 336

    1283 TATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCT 1332
         :::::::::::::::::::::::::|||:::|||:::      :::
     336 hrMetAlaHisGluMetGlyHisAsnPheGlyMetPheHisAspAspTyr 352

    1333 AGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCAT 1382
         |||             |||   |||   ::::::          |||:::||
     353 Ser............CysLysCysProSerThrIle......CysValMe 363

    1383 GGAGGAAACA...GGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCA 1429
         |::::::::::   :::                  ||||||:::||||||
     363 tAspLysAlaLeuSerPheTyrIleProThrAspPheSerSerCysSerA 380

    1430 TTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTC 1479
           |||:::|||      |||::::::          :::   |||||||||
     380 rgLeuSerTyrAspLysPhePheGluAspLysLeuSerAsnCysLeuPhe 396

    1480 AACAGG......CCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGG 1523
         |||          ||||||   :::::::::          |||||||||
     397 AsnAlaProLeuProThrAspIleIleSerThrProIleCysGlyAsnGl 413

    1524 ATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCT 1573
           ||||||   ||||||:::|||||||||||          ||||||
     413 nLeuValGluMetGlyGluAspCysAspCysGlyThrSerGluGluCysT 430

    1574 ATGGATTATGCTGT......AAGAAATGTTCCCTCTCCAACGGGGCTCAC 1617
         ::::::||||||      |||   |||:::::::::          :::
     430 hrAsnIleCysCysAspAlaLysThrCysLysIleLysAlaThrPheGln 446

    1618 TGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACG 1667
         |||:::   |||   |||||||:::       :::|||   |||:::
     447 CysAlaLeuGlyGluCysCysGlu......LysCysGlnPheLysLysAl 461

    1668 AGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATT 1717
          |||      ||||||   |||   :::|||||||||:::   |||   |
     461 aGlyMetValCysArgProAlaLysAspGluCysAspLeuProGluMetC 478

    1718 GTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGA 1767
```

```
     ||:::|||   ||||||:::||||||   :::      :::   :::|||
 478 ysAsnGlyLysSerGlyAsnCysProAspAspArgPheGlnValAsnGly 494

1768 TATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGAC 1817
     :::   |||:::::::::::::|||:::|||      |||   |||   ||
 495 PheProCysHisHisGlyLysGlyHisCysLeuMetGlyThrCysProTh 511

1818 CAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTG 1867
     |   ::::::||||||      :::||||||      :::      :::|
 511 rLeuGlnGluGlnCysThrGluLeuTrpGlyProGlyThrGluValAlaA 528

1868 ACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAAC 1917
     |||||   ||||||::::::   |||      |||::::::   |||
 528 spLysSerCysTyrAsnArg...AsnGluGlyGlySerLysTyrGlyTyr 543

1918 TGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTT 1967
     |||   :::      |||      |||   |||:::   :::|||:::::
 544 CysArgArgValAspAspThrLeuIleProCysLysAlaAsnAspThrMe 560

1968 CTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTC 2017
     :||||||   |||:::|||                             |
 560 tCysGlyLysLeuPheCys.............................G 567

2018 AACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTG 2067
     ||      ::::::   :::|||              :::||||||:::
 567 lnGlyGlySerAspAsnLeuPro............TrpLysGlyArgIle 579

2068 ATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACG......GA 2111
     :::           :::   :::         :::|||||      ::
 580 ValThrPheLeuThrCysLysThrPheAspProGluAspThrSerGlnGl 596

2112 TGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTT 2161
     ::::|||   |||   :::||||||   ||||||   :::   :::|||:
 596 uIleGlyMetValAlaAsnGlyThrLysCysGlyAspAsnLysValCysI 613

2162 TAGATCGGAAGTGCCTACAAATTCAA...GCCCTAAATATGAGCAGCTGT 2208
     :::::   :::|||::::::|||:::   |||   :::   ::::::|||
 613 leAsnAlaGluCysValAspIleGluLysAlaTyrLysSerThrAsnCys 629

2209 CCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAA 2258
          :::||||||         |||:::||||||:::|||||||:::::
 630 ......SerSerLys.........CysLysGlyHisAlaValCysAspHi 641

2259 TGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGC 2304
     :|||      |||   |||:::      |||         ||||||
 641 sGluLeuGlnCysGlnCysGluGluGlyTrpIleProProAspCys 656

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT:AAW90865

seq_documentation_block:
ID   AAW90865 standard; Protein; 775 AA.
XX
AC   AAW90865;
XX
DT   28-JUN-2000  (first entry)
XX
DE   Human ADAM protein #4.
XX
KW   ADAM protein; human; A disintegrin and metalloprotease; diagnosis;
KW   drug development; intervertebral hernia; sciatica; pulmonary fibrosis;
KW   diabetic nephropathy; hepatic fibrosis; glomerulitis; osteopetrosis.
XX
OS   Homo sapiens.
XX
PN   WO200014227-A1.
XX
PD   16-MAR-2000.
XX
PF   02-SEP-1999;   99WO-JP04766.
XX
PR   03-SEP-1998;   98JP-0250115.
XX
PA   (TAKE ) TAKEDA CHEM IND LTD.
XX
PI   Yoshimura K,  Hikichi Y,  Nishimura A;
XX
DR   WPI; 2000-271056/23.
DR   N-PSDB; AAZ89726.
XX
PT   Novel protein belong to A disintegrin and metalloprotease family, with
PT   protease activity and extracellular matrix digesting enzyme activity,
PT   for gene diagnosis and developing drugs for treating e.g. sciatica and
PT   glomerulitis  -
XX
PS   Disclosure; Fig 3-4; 109pp; Japanese.
XX
CC   This invention describes a novel human protein (I) which belongs to the
CC   ADAM (A disintegrin and metalloprotease) protein family. The protein,
CC   peptide fragment and antibody are useful for gene diagnosis and in the
CC   development of drugs to prevent or treat intervertebral hernia, sciatica,
CC   glomerulitis, diabetic nephropathy, hepatic fibrosis, pulmonary fibrosis
CC   or osteopetrosis. This sequence represents a human ADAM protein which is
CC   described in the method of the invention.
XX
SQ   Sequence   775 AA;


alignment_scores:
            Quality:  810.50              Length:    682
              Ratio:  2.026                 Gaps:     26
 Percent Similarity:  58.651   Percent Identity: 31.085

alignment_block:
 US-09-634-252A-2 x AAW90865   ..

 Align seg 1/1  to: AAW90865  from: 1  to: 775

  328 GAAATCACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGA 377
      |||::::::   |||   ||||||
   29 GluValValTyrProIleArgLeu.......................... 36

  378 AAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATA 427
          |||   |||      ::::::|||      :::||||||::::::
   37 .HisProLeuHisLysArgGluAlaLysGluProGluGlnGlnGluGlnP 53

  428 ATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGC 477
        :::             ::::::   ::::::::::::         |||
   53 heGlu............ThrGluLeuLysTyrLysMetThrIleAsnGly 65

  478 TCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGA 527
      :::      :::|||   |||      |||:::::||||||:::
   66 LysIleAlaValLeuTyrLeuLysLysAsnLysAsnLeuLeuAlaProGl 82

  528 TTATGTGGAGATTCACTACGAAAAT...GGGAAA.............C 559
       |||:::|||   :::|||::::::   ||||||             |
   82 yTyrThrGluThrTyrTyrAsnSerThrGlyLysGluIleThrThrSerP 99

  560 CACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGA 609
      |||||               :::   |||||||||:::|||   |||
   99 roGlnIleMet.........AspAspCysTyrTyrGlnGlyHisIleLeu 112

  610 GGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGG 659
      :::   |||   |||   :::::::::|||||||||:::||||||:::||
  113 AsnGluLysValSerAspAlaSerIleSerThrCysArgGlyLeuArgGl 129

  660 CATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGG 709
      |   |||::::::            |||:::||||||||||||:::   :
  129 yTyrPheSerGlnGlyAspGlnArgTyrPheIleGluProLeuSerProI 146

  710 TTCATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAACCTTG 759
      ::|||   :::          |||:::   |||   :::   |||
  146 leHisArgAsp.........GlyGlnGluHisAlaLeuPheLys...... 157

  760 GCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAGGT.. 807
              |||:::   :::   ||||||              |||
  158 .........TyrAsnProAspGluLysAsnTyrAspSerThrCysGlyMe 171
```

```
 808 .GACCAGTGGCCCTTTCTCTCTGAATTACAGTGG.........TTGAAAA 847
 171 tAspGlyValLeuTrpAlaHisAspLeuGlnGlnAsnIleAlaLeuProA 188

 848 GAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAA 897
 188 laThrLysLeuValLysLeuLysAspArgLysValGlnGluHisGluLys 204

 898 TATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCG 947
 205 TyrIleGluTyrTyrLeuValLeuAspAsnGlyGluPheLysArgTyrAs 221

 948 CTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTG 997
 221 nGluAsnGlnAspGluIleArgLysArgValPheGluMetAlaAsnTyrV 238

 998 TGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCT 1047
 238 alAsnMetLeuTyrLys...LysLeuAsnThrHisValAlaLeuValGly 253

1048 GTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGT 1097
 254 MetGluIleTrpThrAspLysAspLysIleLysIleThrProAsnAlaSe 270

1098 GCAGATGCTCCATGAGTTCTCAAAATACCGG............CAGCGCA 1135
 270 rPheThrLeuGluAsnPheSerLysTrpArgGlySerValLeuSerArgA 287

1136 TTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCAC 1185
 287 rgLysArgHis...AspIleAlaGlnLeuIleThrAlaThrGluLeuAla 302

1186 TATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCT...CGCAC 1232
 303 GlyThrThrValGlyLeuAlaPheMetSerThrMetCysSerProTyrSe 319

1233 AAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACAAG 1282
 319 rValGlyValValGlnAspHisSerAspAsnLeuLeuArgValAlaGlyT 336

1283 TATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCT 1332
 336 hrMetAlaHisGluMetGlyHisAsnPheGlyMetPheHisAspAspTyr 352

1333 AGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCAT 1382
 353 Ser............CysLysCysProSerThrIle......CysValMe 363

1383 GGAGGAAACA...GGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCA 1429
 363 tAspLysAlaLeuSerPheTyrIleProThrAspPheSerSerCysSerA 380

1430 TTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTC 1479
 380 rgLeuSerTyrAspLysPhePheGluAspLysLeuSerAsnCysLeuPhe 396

1480 AACAGG......CCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGG 1523
 397 AsnAlaProLeuProThrAspIleIleSerThrProIleCysGlyAsnGl 413

1524 ATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCT 1573
 413 nLeuValGluMetGlyGluAspCysAspCysGlyThrSerGluGluCysT 430

1574 ATGGATTATGCTGT......AAGAAATGTTCCCTCTCCAACGGGGCTCAC 1617
 430 hrAsnIleCysCysAspAlaLysThrCysLysIleLysAlaThrPheGln 446

1618 TGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACG 1667
 447 CysAlaLeuGlyGluCysCysGlu......LysCysGlnPheLysLysAl 461

1668 AGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATT 1717
 461 aGlyMetValCysArgProAlaLysAspGluCysAspLeuProGluMetC 478

1718 GTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGA 1767
 478 ysAsnGlyLysSerGlyAsnCysProAspAspArgPheGlnValAsnGly 494

1768 TATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGAC 1817
 495 PheProCysHisHisGlyLysGlyHisCysLeuMetGlyThrCysProTh 511

1818 CAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTG 1867
 511 rLeuGlnGluGlnCysThrGluLeuTrpGlyProGlyThrGluValAlaA 528

1868 ACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAAC 1917
 528 spLysSerCysTyrAsnArg...AsnGluGlyGlySerLysTyrGlyTyr 543

1918 TGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTT 1967
 544 CysArgArgValAspAspThrLeuIleProCysLysAlaAsnAspThrMe 560

1968 CTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTC 2017
 560 tCysGlyLysLeuPheCysAspGly..................... 568

2018 AACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTG 2067
 569 .....GlySerAspAsnLeuPro...........TrpLysGlyArgIle 579

2068 ATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACG......GA 2111
 580 ValThrPheLeuThrCysLysThrPheAspProGluAspThrSerGlnGl 596

2112 TGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTT 2161
 596 uIleGlyMetValAlaAsnGlyThrLysCysGlyAspAsnLysValCysI 613

2162 TAGATCGGAAGTGCCTACAAATTCAA...GCCCTAAATATGAGCAGCTGT 2208
 613 leAsnAlaGluCysValAspIleGluLysAlaTyrLysSerThrAsnCys 629

2209 CCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAA 2258
 630 ......SerSerLys.........CysLysGlyHisAlaValCysAspHi 641

2259 TGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGC 2304
 641 sGluLeuGlnCysGlnCysGluGluGlyTrpIleProProAspCys 656
```

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT:AAW44120

seq_documentation_block:

```
ID   AAW44120 standard; Protein; 660 AA.
XX
AC   AAW44120;
XX
DT   27-APR-1998  (first entry)
XX
DE   Human ADAM 12 predicted protein sequence.
XX
KW   Human; ADAM 12; reprolysin; inflammatory disorder; interdomainase;
KW   cancer; haemorrhagic disorder; amyloidosis; membrane protein shedding.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..24
FT                   /label= hADAM_12{pro}_domain
FT   Region          54..60
FT                   /label= cysteine_switch
```

```
FT   Domain          229..239
FT                   /label= hADAM_12{met}_domain
FT   Domain          301..340
FT                   /label= hADAM_12{dis}_domain
FT   Domain          574..598
FT                   /label= hADAM_12{egf}_domain
FT   Domain          622..642
FT                   /label= hADAM_12{tmt}_domain
XX
PN   WO9740072-A2.
XX
PD   30-OCT-1997.
XX
PF   16-APR-1997;   97WO-GB01067.
XX
PR   19-APR-1996;   96GB-0008130.
XX
PA   (UYSH-) UNIV SHEFFIELD.
XX
PI   Croucher PI,  Mckie N,  Russell RGG;
XX
DR   WPI; 1997-535779/49.
XX
PT   New isolated ADAM 12 polypeptide(s) - used to develop products for
PT   therapy, prophylaxis or diagnosis of e.g. inflammatory disorders,
PT   cancers, haemorrhagic disorders or amyloidosis
XX
PS   Claim 2; Fig 2; 94pp; English.
XX
CC   The present sequence represents the predicted protein sequence for human
CC   ADAM 12. The present specification also describes products, derived or
CC   produced from ADAM 12, which can be used for regulating, promoting or
CC   reducing membrane protein shedding, membrane protein processing, cell-
CC   cell adhesion/contact/communication (e.g. juxtacrine stimulation),
CC   aggrecan degradation (e.g. via the regulation, promotion or reduction of
CC   aggrecanase activity), degradation of basement membrane components,
CC   osteoclast precursor fusion, collagen type II degradation and/or
CC   removal, or activation of matrix metalloprotease (MMP). They can be used
CC   in the therapy, prophylaxis or diagnosis of disorders or diseases of the
CC   connective tissue, e.g. those requiring aggrecan degradation, regulation
CC   (especially prevention or reduction) of cartilage breakdown or osteo- or
CC   rheumatoid arthritis, or skeletal system disorders or diseases e.g. bone
CC   resorption, post-menopausal osteoporosis, Paget's disease or metastatic
CC   or myeloma associated bone diseases, inflammatory diseases or disorders
CC   e.g. caused by pro-inflammatory cytokines, by cell especially
CC   macrophage, infiltration or rheumatoid arthritis, cancers, especially
CC   for detection and prevention of tumour progression or metastasis,
CC   coagulopathies e.g. thrombo-embolic disorder, haemorrhagic disorders or
CC   diseases or amyloidosis e.g. Alzheimer's disease.
XX
SQ   Sequence   660 AA;


     ent_scores:
          Quality:  774.00              Length:    622
            Ratio:  2.109                 Gaps:     20
 Percent Similarity:  59.003   Percent Identity: 30.707

alignment_block:
 US-09-634-252A-2 x AAW44120   ..

 Align seg 1/1  to: AAW44120  from: 1  to: 660

     634 CTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATGATACCTTCGT 683
         |||:::|||         |||:::|||:::::::::   ::::::
      12 LeuAlaThrValLeuAspLeuArgGlyLeuLeuHisLeuGluAsnAlaSe 28

     684 GTATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACAGGTC 733
          |||   |||||||||||:::::                 |||:::
      28 rTyrGlyIleGluProLeuGlnMet.............SerSerHisP 40

     734 GACCACATATAATC:.....................CAGAAAACCTTG 759
             |||||||||                       :::

      40 heGluHisIleIleTyrArgMetAspAspValTyrLysGluProLeuLys 56

     760 GCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAGGTGA 809
         :::|||      :::|||:::::::::::::   |||   :::     ::
      57 CysGlyValSerAsnLysAspIleGluLysGluThrAlaLysAspGluGl 73

     810 CCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAAGAGAG 859
         ::::   |||         ::::::      |||   ||||||:::|||
      73 uGluGluPro......ProSerMetThrGlnLeuLeuArgArgArgArgA 88

     860 CAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTT 909
         |||||   |||                  :::   :::|||:::||||||
      88 laValLeuPro.................GlnThrArgTyrValGluLeu 98

     910 ATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGC 959
         :::|||||||   |||   :::   |||                :::::::
      99 PheIleValValAspLysGluArgTyrAspMetMetGlyArgAsnGlnTh 115

     960 ACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTT 1009
         :   :::::::::           ::::::|||   :::|||||||:::|
     115 rAlaValArgGluGluMetIleLeuLeuAlaAsnTyrLeuAspSerMetT 132

    1010 ACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGG 1059
         ||      :::||||||   |||:::|||||||||::::::|||   |||
     132 yr...IleMetLeuAsnIleArgIleValLeuValGlyLeuGluIleTrp 147

    1060 ACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCA 1109
         |||:::   :::   |||:::|||:::   :::       :::::|||
     148 ThrAsnGlyAsnLeuIleAsnIleValGlyGlyAlaGlyAspValLeuGl 164

    1110 TGAGTTCTCAAAATACCGGCAGCGCATT.........AAGCAGCATGCTG 1150
          :::|||   ::::::|||:::::::::         :::::::::   |
     164 yAsnPheValGlnTrpArgGluLysPheLeuIleThrArgArgArgHisA 181

    1151 ATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGT 1200
         ||:::::::::|||:::   :::      |||            ::::::
     181 spSerAlaGlnLeuValLeuLysLysGlyPheGly...GlyThrAlaGly 196

    1201 CTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAA 1250
         :::::::::   |||   ||||||||||||:::::::::   |||:::||
     197 MetAlaPheValGlyThrValCysSerArgSerHisAlaGlyGlyIleAs 213

    1251 TGAGTATGGT......CTTCCAATGGCAGTGGCACAAGTATTATCGCAGA 1294
         |   :::|||              :::   |||:::::::::::::::::
     213 nValPheGlyGlnIleThrValGluThrPheAlaSerIleValAlaHisG 230

    1295 GCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCA 1344
         ::|||:::::||||||||:::::   :::   :::
     230 luLeuGlyHisAsnLeuGlyMetAsnHisAspAspGlyArg......... 243

    1345 AAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGG 1394
               ||||||:::   :::      :::||||||||   :::::::||
     244 ......AspCysSerCysGlyAlaLysSerCysIleMet...AsnSerGl 257

    1395 GGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAG 1444
         |:::|||   ||||||:::|||||||:::||||||      :::::::::
     257 yAlaSerGlySerArgAsnPheSerSerCysSerAlaGluAspPheGluL 274

    1445 ACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAAC......AGGCCA 1488
           :::      :::   ||||||   ||||||:::|||      :::|||
     274 ysLeuThrLeuAsnLysGlyGlyAsnCysLeuLeuAsnIleProLysPro 290

    1489 ACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGG 1538
         :::   ::::::      :::|||||||||      |||:::|||||
     291 AspGluAlaTyrSerAlaProSerCysGlyAsnLysLeuValAspAlaGl 307

    1539 GGAGGAGTGTGATTGTGGTTTTCATGTGGAATGC...TATGGATTATGCT 1585
         ||||||||||||||||||         ||||||           ||||
     307 yGluGluCysAspCysGlyThrProLysGluCysGluLeuAspProCysC 324

    1586 GTAAG......AAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGG 1629
         ||:::         |||:::|||::::::   |||:::|||:::   |||
     324 ysGluGlySerThrCysLysLeuLysSerPheAlaGluCysAlaTyrGly 340
```

```
 1630 CCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATG 1679
       ||||||:::::      |||   |||   |||   |||      ||
  341 AspCysCysLysAsp......CysArgPheLeuProGlyGlyThrLeuCy 355

 1680 CCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTACTGGAGACT 1729
      ||||      ::::::||||||||||:::   |||||||||:::|||:::|
  355 sArgGlyLysThrSerGluCysAspValProGluTyrCysAsnGlySerS 372

 1730 CTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATGCATGCAAT 1779
      ||      |||   |||::::::      |||:::||||||   |||:::
  372 erGlnPheCysGlnProAspValPheIleGlnAsnGlyTyrProCysGln 388

 1780 CAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAACCA 1829
      :::|||::::::   ||||||||||||   |||:::      |||   ||
  389 AsnAsnLysAlaTyrCysTyrAsnGlyMetCysGlnTyrTyrAspAlaGl 405

 1830 GTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCT 1879
      |||||||   |||:::|||:::||||||   ::::::   |||   | |:
  405 nCysGlnValIlePheGlySerLysAlaLysAlaAlaProLysAspCysP 422

 1880 ATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGAT 1929
      ::   ::::::|||::::::|||   :::   ||||||||||||   :::
  422 heIleGluValAsnSerLysGlyAspArgPheGlyAsnCysGlyPheSer 438

 1930 GGAGACCGGTGG..................................... 1941
      |||:::::::::
  439 GlyAsnGluTyrLysLysCysAlaThrGlyLeuSerLeuLysPheHisAl 455

 1941 ..... 1941

  455 aProPheLeuSerThrMetLeuGlnGluAlaValArgGlnThrGlyThrT 472

 1942 ..............ATTCAGTGCAGCAAACATGAT..... 1962
                     :::   |||   |||   |||
  472 yrLeuGlyGlySerValCysCysMetLysSerAspCysArgIleValThr 488

 1963 ............GTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCG 2000
                  ::::::||||||   |||   |||   |||:::   ::
  489 LeuValLysAsnAlaLeuCysGlyLysLeuGlnCysGluAsnValGlnGl 505

 2001 AGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCT 2050
      :   |||   :::|||                  |||:::|||:::   :::
  505 uIleProValPheGly.............IleValProAlaIleIleG 517

 2051 ACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGAT 2100
            :::|||         |||   |||:::      |||
  517 lnThrProSerArgGlyThrLysCysTrpGlyValAspPheGlnLeuGly 533

 2101 GATGATACG...GATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCC 2147
      :::|||:::   |||   |||   |||:::::::||||||   ||||||
  534 SerAspValProAspProGlyMetValAsnGluGlyThrLysCysGlyAl 550

 2148 GTCTATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATA 2197
       :::   :::|||   :::   :::|||::::::   ::::::||||||
  550 aGlyLysIleCysArgAsnPheGlnCysValAspAlaSerValLeuAsnT 567

 2198 TGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGG 2247
            :::|||   :::::::::|||          |||   |||||||||
  567 yr...AspCysAspValGlnLysLys.........CysHisGlyHisGly 579

 2248 GTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGAC 2297
      ||||||:::::::::   :::|||   |||:::      ||||||
  580 ValCysAsnSerAsnLysAsnCysHisCysGluAsnGlyTrpAlaProPr 596

 2298 AGATTGCAGTATCCGG 2313
       :::|||:::   :::
  596 oAsnCysGluThrLys 601

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW73009

seq_documentation_block:
ID   AAW73009 standard; Protein; 613 AA.
XX
AC   AAW73009;
XX
DT   02-FEB-1999  (first entry)
XX
DE   Cobra venom protease mocarhagin NMM-9.
XX
KW   Mocarhagin; snake venom; Mozambiquan spitting cobra; protease;
KW   inflammation; myocardial infarction; thrombosis; infection;
KW   metastasis; therapy; NMM-9.
XX
OS   Naja mossambica mossambica.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..23
FT                   /label= Sig_peptide
FT   Protein         24..613
FT                   /label= Mat_protein
XX
PN   WO9846771-A2.
XX
PD   22-OCT-1998.
XX
PF   14-APR-1998;   98WO-US07998.
XX
PR   18-FEB-1998;   98US-0026001.
PR   15-APR-1997;   97US-0843373.
PR   23-JAN-1998;   98US-0012637.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Boodhoo A,  Sako D,  Seehra JS,  Shaw G;
XX
DR   WPI; 1998-568735/48.
DR   N-PSDB; AAV07897.
XX
PT   Isolated mocarhagin cobra venom protease, and nucleic acids encoding
PT   it - used to develop products for treating e.g. myocardial
PT   infarction, thrombosis, bacterial or viral infection, metastatic
PT   conditions or inflammatory disorders
XX
PS   Claim 40; Page 45-48; 97pp; English.
XX
CC   This is the amino acid sequence of mocarhagin NMM-9, a highly
CC   specific metalloproteinase from the venom of the Mozambiquan
CC   spitting cobra.  The invention provides mocarhagin polypeptides
CC   (see AAW73007-13) and polynucleotides (see AAV07895-901) encoding them,
CC   as well as host cells and methods of producing the (especially
CC   mature) polypeptides.  Mocarhagin proteins are capable of cleaving
CC   anionic polypeptide containing sulphated tyrosine residues,
CC   P-selectin glycoprotein (GP) ligand-1 (PSGL-1) and GP1b-alpha
CC   (claimed).  They also inhibit neutrophil/HL60 binding, inhibit
CC   platelet binding to von Willebrand Factor, require Ca2+ and Zn2+
CC   ions for activity and have activity inhibited by excess EDTA or
CC   high concentrations of DFP (claimed).  They can be used to inhibit
CC   selectin-mediated binding and to treat inflammatory disease
CC   (claimed).  In particular, they can be used to treat e.g. myocardial
CC   infarction, vessel restenosis, thrombosis, bacterial or viral
CC   infection, metastatic conditions, inflammatory disorders such as
CC   arthritis, acute respiratory distress syndrome, asthma, emphysema,
CC   delayed type hypersensitivity reaction, systemic lupus
CC   erythematosus, thermal injury such as burns or frostbite,
CC   autoimmune thyroiditis, experimental allergic encephalomyelitis,
CC   multiple sclerosis, multiple organ injury syndrome secondary to
CC   trauma, diabetes, Reynaud's syndrome, neutrophilic dermatosis
CC   (Sweet's syndrome), inflammatory bowel disease, Grave's disease,
CC   glomerulonephritis, gingivitis, periodontitis, haemolytic uremic
CC   syndrome, ulcerative colitis, Crohn's disease, necrotising
CC   enterocolitis, granulocyte transfusion associated syndrome,
CC   cytokine-induced enterocolitis, granulocyte transfusion associated
CC   syndrome, or cytokine-induced toxicity.  Mocarhagin protein may
CC   also be useful in organ transplantation, both to prepare organs for
CC   transplantation and to quell organ transplant rejection, to treat
```

```
CC   haemodialysis and leukophoresis patients, or as an inhibitor of P-
CC   or E-selectin-mediated intercellular adhesion.
XX
SQ   Sequence   613 AA;


alignment_scores:
            Quality:  772.50              Length:    657
              Ratio:  1.991                 Gaps:     21
 Percent Similarity:  59.056   Percent Identity: 28.311

alignment_block:
 US-09-634-252A-2 x AAW73009   ..

 Align seg 1/1  to: AAW73009  from: 1  to: 613

     283 CAGAATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAAT 332
         |||:::||||||      ::::::   :::|||:::   :::   |||::
      17 GlnGlySerSerIleIleLeuGluSerGlyAsnValAsnAspTyrGluVa 33

     333 CACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGAAAGCC 382
         ::::   |||:::::::::          :::          ::::::|
      33 lValTyrProGlnLysValProAlaLeuSerLysGlyGlyValGlnAsnP 50

     383 CTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAG 432
         ||                                  |||   :::::|||
      50 ro..............................GlnProGluThrLys 55

     433 GCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAA 482
                       |||   :::||||||::::::      |||:::
      56 TyrGluAspThrMetGlnTyrGluPheGlnValAsn......GlyGluPr 70

     483 ATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATG 532
         :::|||   |||      |||:::|||||||:::|||:::||||||:
      70 oValValLeuHisLeuGluArgAsnLysGlyLeuPheSerGluAspTyrT 87

     533 TGGAGATTCACTACGAAAATGGGAAACCACAGTACTCTAAGGGT...... 576
         ::|||   ||||||   :::        :::   :::   :::
      87 hrGluThrHisTyrAlaSerAspGlyArgGluIleThrThrSerProLeu 103

     577 ...GGAGAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTC 623
            :::||||||||||||||||||   |||:::::      |||||
     104 ValGlnAspHisCysTyrTyrHisGlyTyrIleGlnAsnGluAlaAspSe 120

     624 CAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATG 673
         |::::::::::::|||:::|||:::||||||   |||   ||||||   :
     120 rSerAlaValIleSerAlaCysAspGlyLeuLysGlyHisPheGluLeuG 137

     674 ATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAA 723
         ::      :::|||:::||||||||||||:::::
     137 lnGlyGluThrTyrPheIleGluProLeuLysIle.............. 148

     724 AGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCAGGACAGTATTC 773
         |||   ::::::   |||   |||
     149 SerAspSerGluAlaHisAlaIle........................... 156

     774 TAAGCAAATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCC.... 819
                 |||:::   ::::::|||:::   |||:::   |||
     157 .......TyrLysAspGluAsnValGluAsnGluAspGluThrProGluT 171

     820 .....TTTCTCTCTGAATTACAGTGG.............TTGAAAAGA 849
                ::::::|||      |||               :::::::::
     171 hrCysGlyValThrGluThrThrTrpGluSerAspGluSerIleGluLys 187

     850 AGGAAGAGAGCAGTGAAT...CCATCACGTGGTATATTT...GAAGAAAT 893
            ::::::   :::|||   |||::::::      :::   :::
     188 ThrSerGlnLeuThrAsnThrProGluGlnAspArgTyrLeuGlnAlaLy 204

     894 GAAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGC 943
          ||||||||||||::   :::|||   |||:::      |||:::   :
     204 sLysTyrLeuGluPheTyrValValValAspAsnIleMetTyrArgHisT 221

     944 ATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAAC 993

         :::::   :::           ::::::   :::   :::::::::|||
     221 yrLysArgAspLysProValIleLysArgArgValTyrGluMetIleAsn 237

     994 CTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGT 1043
         ::::::   :::|||   ::::::||||||   ::::::::::|||::
     238 ThrMetAsnMetValTyr...AsnArgLeuAsnPheHisIleAlaLeuIl 253

    1044 GGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACC 1093
         :::::::|||   |||:::::::::::::::|||::::::   ::::::
     253 eGlyLeuGluIleTrpSerAsnArgAsnGluIleAsnValGlnSerAspV 270

    1094 CTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGC......... 1134
                    |||       |||::::::::::|||::::::
     270 alGlnAlaThrLeuAspLeuPheGlyGluTrpArgGluLysLysLeuLeu 286

    1135 ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCA 1184
         :::::::::   |||   :::::::|||::::::   :::   |||
     287 ProArgLysArgAsnAspAsnAlaGlnLeuLeuThrGlyIleAspPheLy 303

    1185 CTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAA 1234
                    :::|||:::|||:::|||:::::::|||:::
     303 sGlyThrProValGlyLeuAlaTyrIleGlySerIleCysAsnProLysS 320

    1235 GAGGAGTTGGT...GTGAATGAGTATGGTCTTCCAATGGCA...GTGGCA 1278
           :::|||:::   |||:::::::|||:::          :::   ||||||
     320 erSerValAlaValValGlnAspTyrSerSerArgThrSerMetValAla 336

    1279 CAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACC 1328
         :::::::::::::::::::::::|||:::|||||||:::   :::
     337 IleThrMetAlaHisGluMetGlyHisAsnMetGlyIleHisHisAsp.. 352

    1329 TTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCA 1378
                    |||:::|||   |||   ::::::          |||:
     353 ..........GlyProSerCysThrCysGlySerAsn......LysCysV 364

    1379 TCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGC 1428
         ::|||:::        ::::::   :::   :::||||||:::||||||
     364 alMetSerThrArgArgThrGluProAlaTyrGlnPheSerSerCysSer 380

    1429 ATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTT 1478
         :::   |||:::::::::::|||   |||            |||:::::
     381 ValArgGluHisGlnGluTyrLeuLeuArgAspArgProGlnCysIleLe 397

    1479 CAACAGGCCA......ACAAAGCTATTTGAGCCCACGGAATGTGGAAATG 1522
         :|||:::|||      |||   :::   :::|||      |||||||||:
     397 uAsnLysProLeuSerThrAspIleValSerProProIleCysGlyAsnA 414

    1523 GATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGC 1572
         :::::||||||:::||||||||||||||||||||      ::::::|||
     414 snPheValGluValGlyGluGluCysAspCysGlySerProAlaAspCys 430

    1573 TATGGATTATGCTGT......AAGAAATGTTCCCTCTCCAACGGGGCTCA 1616
            :::   ||||||            |||:::|||:::      |||::
     431 GlnSerAlaCysCysAspAlaThrThrCysLysLeuGlnProHisAlaGl 447

    1617 CTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCAC 1666
         :|||::::::      |||||||:::      :::|||   |||:::
     447 nCysAspSerGluGlyCysCysGlu......LysCysLysPheLysGlyA 462

    1667 GAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATAT 1716
           |||   |||||||||   |||   ::::::||||||:::   |||
     462 laGlyAlaGluCysArgAlaAlaLysAspAspCysAspLeuProGluLeu 478

    1717 TGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGG 1766
         |||||||||:::|||:::::::||||||   ::::::   :::::::::||
     479 CysThrGlyGlnSerAlaGluCysProThrAspIlePheGlnArgAsnGl 495

    1767 ATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGA 1816
         |      |||::::::|||:::|||   ||||||||||||:::|||
     495 yLeuProCysGlnAsnAsnGluGlyTyrCysTyrAsnGlyLysCysProI 512

    1817 CCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCT 1866
                |||||||||      :::   |||      :::      |||
```

```
     512 leMetThrAsnGlnCysIleAlaLeuArgGlyProGlyValLysValSer 528

    1867 GACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAA 1916
                ||:::      ||||||   :::   :::      ||
     529 ArgAspSerCysPhe...ThrLeuAsnGlnArgThrSerGlyCysGlyLe 544

    1917 CTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGT 1966
          |||      :::      |||   |||   |||:::      |||| |
     544 uCysArgMetGluTyrGlyArgLysIleProCysAlaAlaLysAspValL 561

    1967 TCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGT 2016
           ||||||   |||:::|||   :::
     561 ysCysGlyArgLeuPheCysLysLys..................... 569

    2017 CAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGT 2066
                                                 ::::::   ::
     570 .................................GlyAsnSerMetIl 574

    2067 GATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGG 2116
          :   :::||||||           |||         |||   :::   |
     574 eCysAsnCysSer...........ValSerProArgAspProSerTyrG 587

    2117 GCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGAT 2166
          ||   ||||||   ||||||   ||||||   :::|||:::|||   :::
     587 lyMetValGluProGlyThrLysCysGlyAspGlyMetValCysSerAsn 603

    2167 CGGAAGTGCCTACAAATTCAA 2187
         |||:::|||:::::::::::
     604 ArgGlnCysValAspValLys 610

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW73013

seq_documentation_block:
ID   AAW73013 standard; Protein; 621 AA.
XX
AC   AAW73013;
XX
DT   02-FEB-1999  (first entry)
XX
DE   Cobra venom mocarhagin NMM-9ek.
XX
KW   Mocarhagin; snake venom; Mozambiquan spitting cobra; protease;
KW   inflammation; myocardial infarction; thrombosis; infection;
KW   metastasis; therapy; NMM-9ek.
XX
OS   Naja mossambica mossambica.
     Synthetic.

     Key             Location/Qualifiers
FT   Peptide         1..196
FT                   /label= Pro_peptide
FT   Cleavage-site   192..196
FT                   /note= "enterokinase cleavage site"
FT   Protein         197..621
FT                   /label= Mat_protein
XX
PN   WO9846771-A2.
XX
PD   22-OCT-1998.
XX
PF   14-APR-1998;   98WO-US07998.
XX
PR   18-FEB-1998;   98US-0026001.
PR   15-APR-1997;   97US-0843373.
PR   23-JAN-1998;   98US-0012637.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Boodhoo A,  Sako D,  Seehra JS,  Shaw G;
XX
DR   WPI; 1998-568735/48.
DR   N-PSDB; AAV07901.
XX
PT   Isolated mocarhagin cobra venom protease, and nucleic acids encoding
PT   it - used to develop products for treating e.g. myocardial
PT   infarction, thrombosis, bacterial or viral infection, metastatic
PT   conditions or inflammatory disorders
XX
PS   Claim 80; Page 65-68; 97pp; English.
XX
CC   This is the amino acid sequence of a modified cobra venom mocarhagin
CC   protein, termed NMM-9ek, that includes an enterokinase cleavage site
CC   between the propeptide and mature peptide of mocarhagin (see also
CC   AAW73009).  Introduction of the cleavage site may allow secretion of
CC   active mocarhagin from eukaryotic host cells.  The invention
CC   provides mocarhagin polypeptides (see AAW73007-13) and polynucleotides
CC   (see AAV07895-901), as well as host cells and methods of producing
CC   (especially mature) polypeptides.  Mocarhagin proteins are capable
CC   of cleaving anionic polypeptide containing sulphated tyrosine
CC   residues, P-selectin glycoprotein (GP) ligand-1 (PSGL-1) and
CC   GP1b-alpha (claimed).  They also inhibit neutrophil/HL60 binding,
CC   inhibit platelet binding to von Willebrand Factor, require Ca2+ and
CC   Zn2+ ions for activity and have activity inhibited by excess EDTA
CC   or high concentrations of DFP (claimed).  They can be used for
CC   inhibiting selectin-mediated binding and for treating an
CC   inflammatory disease (claimed).  In particular, they can be used
CC   for treating e.g. myocardial infarction, vessel restenosis,
CC   thrombosis, bacterial or viral infection, metastatic conditions,
CC   inflammatory disorders such as arthritis, acute respiratory distress
CC   syndrome, asthma, emphysema, delayed type hypersensitivity reaction,
CC   systemic lupus erythematosus, thermal injury e.g. burns or frostbite,
CC   autoimmune thyroiditis, experimental allergic encephalomyelitis,
CC   multiple sclerosis, multiple organ injury syndrome secondary to
CC   trauma, diabetes, Reynaud's syndrome, neutrophilic dermatosis
CC   (Sweet's syndrome), inflammatory bowel disease, Grave's disease,
CC   glomerulonephritis, gingivitis, periodontitis, haemolytic uremic
CC   syndrome, ulcerative colitis, Crohn's disease, necrotising
CC   enterocolitis, granulocyte transfusion associated syndrome,
CC   cytokine-induced enterocolitis, granulocyte transfusion associated
CC   syndrome, or cytokine-induced toxicity.  Mocarhagin protein may
CC   also be useful in organ transplantation, both to prepare organs for
CC   transplantation and to quell organ transplant rejection, to treat
CC   haemodialysis and leukophoresis patients, or as an inhibitor of P-
CC   or E-selectin-mediated intercellular adhesion.
XX
SQ   Sequence   621 AA;


alignment_scores:
            Quality:  766.50              Length:    660
              Ratio:  1.981                 Gaps:     19
 Percent Similarity:  58.636   Percent Identity: 28.182

alignment_block:
 US-09-634-252A-2 x AAW73013   ..

 Align seg 1/1  to: AAW73013  from: 1  to: 621

     283 CAGAATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAAT 332
         |||:::||||||      ::::::   :::|||:::   :::   |||::
      17 GlnGlySerSerIleIleLeuGluSerGlyAsnValAsnAspTyrGluVa 33

     333 CACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGAAAGCC 382
         ::::   |||:::::::::           :::          ::::::|
      33 lValTyrProGlnLysValProAlaLeuSerLysGlyGlyValGlnAsnP 50

     383 CTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAG 432
         ||                                |||   ::::::|||
      50 ro..............................GlnProGluThrLys 55

     433 GCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAA 482
                       |||   :::||||||::::::      |||:::
      56 TyrGluAspThrMetGlnTyrGluPheGlnValAsn......GlyGluPr 70

     483 ATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATG 532
             :::|||   |||      |||:::|||||||:::|||:::|||||||:
```



seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT:AAR99801

seq_documentation_block:
ID AAR99801 standard; Protein; 709 AA.

```
XX
AC   AAR99801;
XX
DT   02-JAN-1997  (first entry)
XX
DE   CRII-7 nerve protein, facilitates regeneration of nerve cells.
XX
KW   Probe; central nervous system; peripheral nervous system; CNS; PNS;
KW   axon; neurone; neuronal cells; glial cells; Schwann cells; trauma;
KW   pathology; regeneration; MS2; metalloprotease.
XX
OS   Homo sapiens.
XX
PN   WO9617865-A2.
XX
PD   13-JUN-1996.
XX
PF   05-DEC-1995;   95WO-EP04777.
XX
PR   27-JAN-1995;   95DE-4002525.
PR   05-DEC-1994;   94DE-4443159.
XX
PA   (BOEF ) BOEHRINGER MANNHEIM GMBH.
XX
PI   Gillen C,  Gleichmann M,  Mueller H;
XX
DR   WPI; 1996-287114/29.
DR   N-PSDB; AAT34616.
XX
PT   Nucleic acids associated with damaged or regenerating nerve cells -
PT   for diagnostic or therapeutic use.
XX
PS   Claim 4; Page 50-52; 54pp; German.
XX
CC   Nucleic acids described in AAT34587-T34616 are useful as probes for
CC   detecting traumatic and pathological changes in the peripheral
CC   nervous system.  They may also be used to identify hybridising
CC   sequences from an ischial nerve cDNA library, these sequences encode
CC   proteins expressed in damaged and/or regenerating nerve cells. They
CC   can be used to transfect cells which then express the protein which
CC   can then be harvested and studied.  The nucleic acids are preferably
CC   selected from NTII-11 which corresponds to plasmolipin cDNA with a
CC   234 bp N-terminal extension and is expressed in glial cells of the
CC   central nervous system and Schwann cells of the peripheral nervous
CC   system; NTII-1, which codes for a protein having 79% homology with
CC   human CDC4L in three domains and may be involved in regulating
CC   Schwann cell growth as well as being therapeutically useful for
CC   regenerating nerve cells; and CRII-7 which codes for this protein
CC   which is similar to the macrophage-specific cysteine-rich mouse
CC   protein MS2 and a metalloprotease and which inhibits interaction
CC   between Schwann cells and the extracellular matrix and is useful
CC   for the therapy of nerve damage by facilitating regeneration of
CC   damaged axons.
XX
SQ   Sequence   709 AA;


alignment_scores:
            Quality:  765.50              Length:    674
              Ratio:  1.919                 Gaps:     25
 Percent Similarity:  59.199   Percent Identity: 29.377

alignment_block:
 US-09-634-252A-2 x AAR99801   ..

 Align seg 1/1  to: AAR99801  from: 1  to: 709

     580 GAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGT 629
         |||:::|||   |||:::|||   :::::|||      :::|||   ||
       9 GluAsnCysCysTyrArgGlyArgValGlnGlyArgProSerSerTrpVa 25

     630 GGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTT.........GAAG 670
         |:::|||   :::|||:::|||:::::::|||:::              :

      25 lSerLeuCysAlaCysSerGlyIleArgGlyLeuValValLeuSerProG 42

     671 ATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAG 720
         ::   :::::::::      :::   |||   :::|||
      42 luArgSerTyrThrLeuGluLeuGlyProGlyAspLeu............ 54

     721 AAAAGCACAGGTCGACCACATATAATCCAGAAA............ACCTT 758
                   ||||||   |||:::::::::                  ||
      55 .........GlnArgProLeuIleValSerArgIleGlnAspLeuLeuLe 68

     759 GGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAGGTG 808
         |   |||:::                  :::      :::
      68 uProGlyHisThrCysAlaProSerTrpHisAlaPheValProThrGluA 85

     809 ACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAAGAGA 858
                 :::|||   |||   :::   |||:::|||   |||
      85 laAlaProAspLeuLeuLeuGluGlnHisHisLeuArgArgLeuLys... 100

     859 GCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACT 908
                     |||   :::      |||   |||   :::|||||
     101 ..............ArgAspValValThrGluThrLysIleValGluLe 112

     909 TATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATG 958
         |:::||||||   |||:::::      :::|||:::   :::   :::
     112 uValIleValAlaAspAsnSerGluValArgLysTyrProAspPheGlnG 129

     959 CACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATT 1008
               |||:::   :::   :::|||:::   |||:::|||::::::
     129 lnLeuLeuAsnArg...ThrLeuGluValAlaLeuLeuLeuAspThrPhe 144

    1009 TACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTG 1058
         :::   :::   ||||||:::||||||:::||||||::::::|||:::||
     145 Phe...GlnProLeuAsnValArgValAlaLeuValGlyLeuGluAlaTr 160

    1059 GACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCC 1108
         ||||:::::::|||   |||:::::::::::::||||||:::   :::|||
     160 pThrGlnArgAspLeuIleGluMetSerSerAsnProAlaValLeuLeuA 177

    1109 ATGAGTTCTCAAAATACCGGCAG..............CGCATTAAGCAG 1143
           :::|||   ::::::|||:::               |||:::   :::
     177 spAsnPheLeuArgTrpArgArgThrAspLeuLeuProArgLeuProHis 193

    1144 CATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAG 1193
               |||:::::::::|||::::::   :::::::|||
     194 ......AspSerAlaGlnLeuValThrValThrSerPheSerGlyProMe 208

    1194 AAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGC......ACAAGAG 1237
             :::::::::      :::::::::||||||         :::   |
     208 tValGlyMetAlaIleGlnAsnSerIleCysSerProAspPheSerGlyG 225

    1238 GAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACAAGTATTA 1287
         |||||:::::::::::::              ::::::||||||:::   :::
     225 lyValAsnMetAspHisSerThrSerIleLeuGlyValAlaSerSerIle 241

    1288 TCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAG 1337
         :::::::::|||::::::::::||||||::::::   :::      :::::
     242 AlaHisGluLeuGlyHisSerLeuGlyLeuAspHisAspSerProGlyAs 258

    1338 AAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCATGGAG. 1386
         ::::      |||                      :::||||||||||||
     258 nSerCysProCysProGlyProAlaProAlaLysSerCysIleMetGluA 275

    1387 ..GAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTTG 1434
           :::|||            :::   :::||||||:::||||||
     275 laSerThrAspPheLeuProGlyLeuAsnPheSerAsnCysSerArgTrp 291

    1435 GAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAAC.. 1482
             ::      |||      |||   |||:::||||||||:::
     292 AlaLeuGluLysAlaLeuLeuAspGlyMetGlySerCysLeuPheGluTr 308

    1483 .AGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATACGTGG 1531
           |||::::::            :::   |||||||||   :::|||:
     308 pProProSerArgAlaProMetSerSerLeuCysGlyAsnMetPheValA 325
```

```
1532 AAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTATGGATTA 1581
     ::   ||||||:::|||||||||||||||||      ||||||
 325 spProGlyGluGlnCysAspCysGlyPheProAspGluCysThrAspPro 341

1582 TGCTGTAAG......AAATGTTCCCTCTCCAACGGGGCTCACTGC...AG 1622
     ||||||           |||:::|||      ||||||:::|||   ||
 342 CysCysAspTyrPheThrCysGlnLeuArgProGlyAlaGlnCysAlaSe 358

1623 CGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGT 1672
     ||||||||||||||||:::|||      |||   :::||||||   |||:
 358 rAspGlyProCysCysGlnAsn......CysLysLeuGlnProAlaGlyT 373

1673 ATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTACT 1722
     :::::||||||      :::::::::||||||:::   |||:::|||
 373 rpGlnCysArgLeuProThrAspAspCysAspLeuProGluPheCysLeu 389

1723 GGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATGC 1772
     |||||||||:::|||||||||||||:::::::::      ||||||
 390 GlyAspSerSerGlnCysProProAspIleArgLeuGlyAspGlyGluPr 406

1773 ATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGAG 1822
      |||   ::::::::::::   |||   :::|||:::|||   :::
 406 oCysAlaSerGlyGluAlaValCysMetHisGlyArgCysAlaSerTyrT 423

1823 ACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAAG 1872
       :::|||||||||   :::||||||      |||      :::
 423 hrArgGlnCysGlnSerLeuTrpGlyProGlyAlaGlnProAlaAlaPro 439

1873 TTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCGG 1922
     :::|||   :::      ||||||:::|||:::      |||:::|||||
 440 LeuCysLeuGlnThrAlaAsnThrArgGlyAsnAlaPheGlySerCysGl 456

1923 GAAGGAT...GGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCT 1969
     |:::::::   :::      ::::::   |||:::   :::|||:::::|
 456 yArgSerProSerGlySerTyrMetProCysAsnLeuArgAspAlaIleC 473

1970 GTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCAA 2019
     |||||   |||   |||            :::   |||   :::|||:::
 473 ysGlyGlnLeuGlnCysGlnTrpGlyArgAsnGlnProLeuLeuGlySer 489

2020 CTTCAG.........GGTGAGATCATTCCAACTTCCTTCTACCATCAAGG 2060
     :::|||         :::|||::::::   ::::::             ||
 490 ValGlnAspGlnLeuSerGluValLeuGluAlaAsn............Gl 502

2061 CCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACGG 2110
     |      ::::::||||||         |||   ||||||      :::|
 502 yThrGlnLeuAsnCysSer......TrpValAspLeuAspLeuGlyAsnA 517

2111 ATGTGGGC.........TATGTAGAAGATGGAACGCCATGTGGCCCGTCT 2151
     |||||:::          :::      ||||||   ||||||||||:::
 517 spValAlaGlnProLeuLeuAlaLeuProGlyThrAlaCysGlyProGly 533

2152 ATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAG 2201
     ::::::|||:::   ::::::|||      ::::::   |||:::   ::
 534 LeuValCysIleGlyHisArgCysGlnProValAspLeuLeuGlyAlaGl 550

2202 CAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGT 2251
     ::::|||           ::::::|||   |||   ||||||||||||
 550 nGluCys...........ArgSerLys...CysHisGlyHisGlyValC 562

2252 GTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGAT 2301
     ||:::::::::      |||   ||||||      ||||||      |||
 562 ysAspSerSerArgHisCysHisCysAspGluGlyTrpAlaProProAsp 578

2302 TGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAAGGATGAAGG 2351
     |||      :::      :::|||
 579 CysMetThrGln......LeuArg..................... 584

2352 ACCCAAGGGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGGTG 2401
                :::::::::::|||      |||   :::   :::
 585 ............AlaThrSerSerLeuThrThrGlyLeuLeuSerL 597
```

```
2402 CCATCCTGGTAGCAGCTATTGTC...CTTGGGGGCACAGGCTGGGGATTT 2448
          :::|||:::   ::::::|||   ||||||::::::   |||
 597 euLeuLeuLeuLeuValLeuValLeuLeuGlyAlaSerTyrTrpTyrArg 613

2449 AAAAATGTCAAGAAGAGA.........................AGGTT 2471
          ::::::   :::|||                          :::::
 614 AlaArgLeuHisGlnArgLeuCysGlnLeuLysGlySerSerCysGlnTy 630

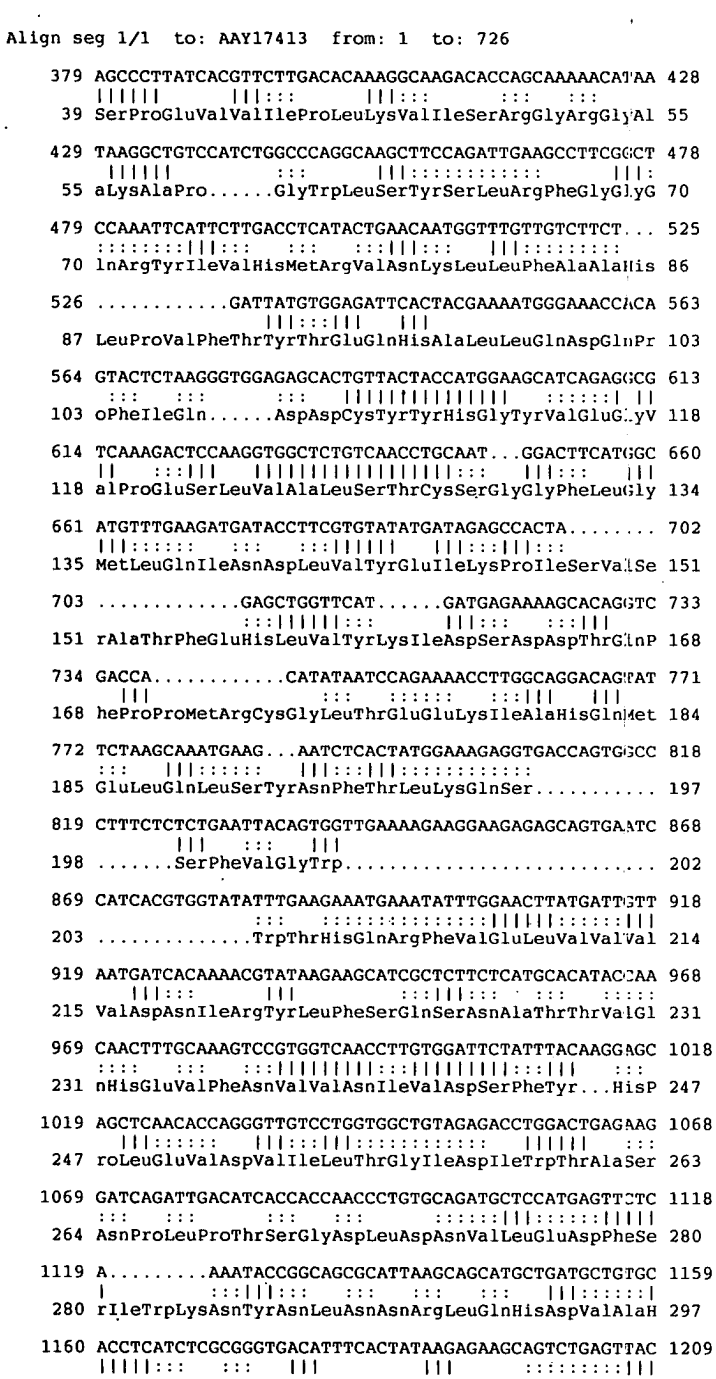
2472 CGATCCTACTCAGCAAGGCCCC 2493
     :      :::|||:::||||||
 630 rArgAlaAlaGlnSerGlyPro 637
```

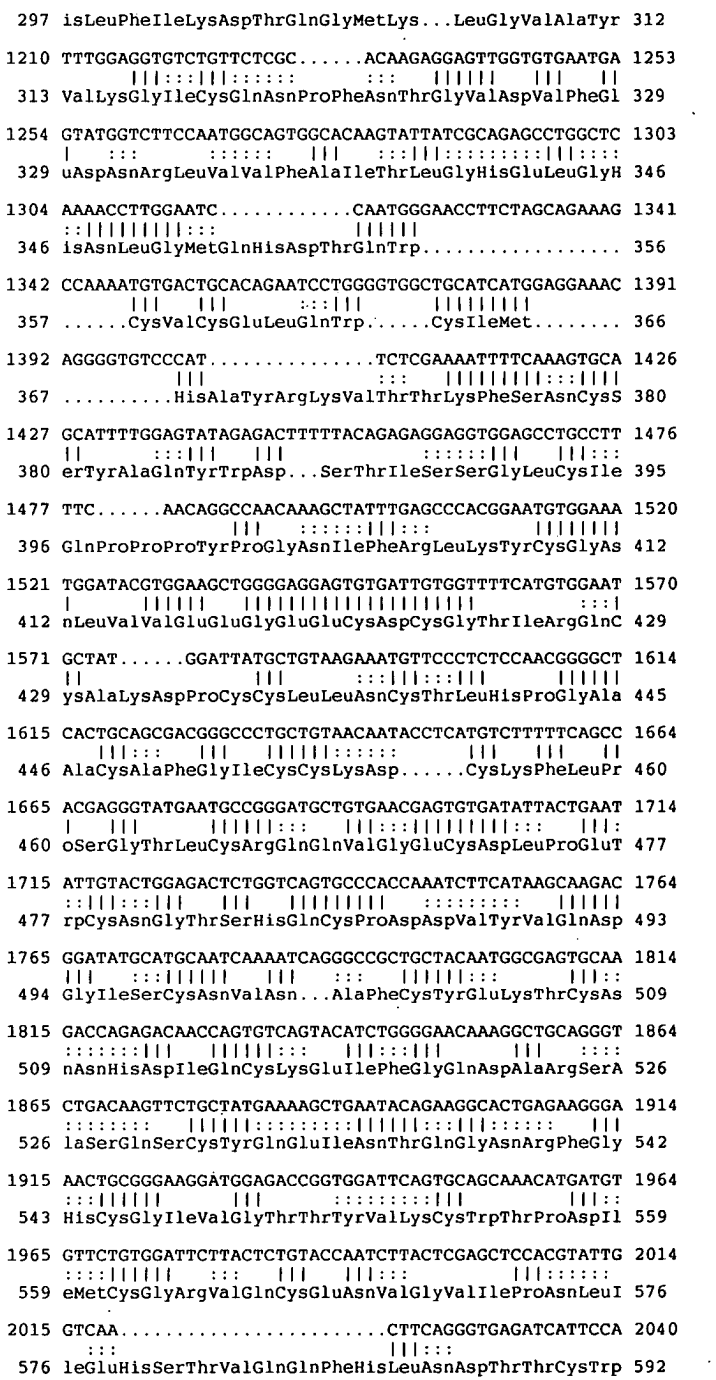
seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY17413

seq_documentation_block:
ID AAY17413 standard; Protein; 726 AA.
XX
AC AAY17413;
XX
DT 26-JUL-1999 (first entry)
XX
DE Human SVPH1-26 protein.
XX
KW Human; SVPH1-26; proteinase; testis; fertilisation; spermatogenesis;
KW birth control; detergent additive; diagnosis; testicular cancer.
XX
OS Homo sapiens.
XX
PN WO9923228-A1.
XX
PD 14-MAY-1999.
XX
PF 30-OCT-1998; 98WO-US22965.
XX
PR 30-OCT-1997; 97US-0063571.
XX
PA (IMMV) IMMUNEX CORP.
XX
PI Cerretti DP;
XX
DR WPI; 1999-337487/28.
DR N-PSDB; AAX56461.
XX
PT New human SVPH1-26 DNA useful for the diagnosis and prognosis of
PT testicular cancers
XX
PS Claim 2; Page 73-83; 96pp; English.
XX
CC The present sequence is human SVPH1-26. Human SVPH1-26 polypeptides
CC are proteinases implicated in fertilisation and spermatogenesis. They
CC can be used as therapeutic agents. A proteinase inhibitor of the
CC catalytic domain would inhibit SVPH1-26 activity and would be useful as
CC a method for birth control. Also, an inhibitor of the disintegrin domain
CC of SVPH1-26 may affect fertilisation. The proteinase activity of SVPH1-26
CC can also be used as a detergent additive for the removal of stains having
CC a protein component. The SVPH1-26 polypeptides and fragments can also be
CC used as molecular weight markers, as markers for determination of
CC isoelectric points of sample proteins and as controls for establishing
CC the extent of fragmentation of a protein sample. The products can also be
CC used for identifying, separating or purifying cells that express SVPH1-26
CC polypeptides such as testis cells. They can be used for the diagnosis
CC and prognosis of testicular cancers.
XX
SQ Sequence 726 AA;

alignment_scores:
Quality: 765.00 Length: 721
Ratio: 1.898 Gaps: 29
Percent Similarity: 55.895 Percent Identity: 29.542

alignment_block:
US-09-634-252A-2 x AAY17413 ..


Align seg 1/1 to: AAY17413 from: 1 to: 726



```
    2041 ACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGT 2090
           :::   ||||||   |||   :::|||
     593 GlyThrAspTyrHisLeuGlyMetAlaIle..................... 602

    2091 AGTTTTAGATGATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCAT 2140
                        |||:::|||   |||:::|||||||||   |
     603 ................ProAspIleGlyGluValLysAspGlyThrValC 614

    2141 GTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCC 2190
         |||||||||:::   :::|||:::   :::||||||   ::::::
     614 ysGlyProGluLysIleCysIleArgLysLysCysAlaSerMetValHis 630

    2191 CTAAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGG 2240
         |||::::::::::      |||               |||:::|||:::
     631 LeuSerGlnAlaCysGlnPro..............LysThrCysAsnMe 642

    2241 CCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGG 2290
          :::|||:::|||:::|||:::      |||   |||:::       ||||
     642 tArgGlyIleCysAsnAsnLysGlnHisCysHisCysAsnHisGluTrpA 659

    2291 CAGGGACAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCC 2340
         ||        |||
     659 laProProTyrCys..................................... 663

    2341 AAGGATGAAGGA..............CCCAAGGGTCCTAGTGCCACCAA 2375
         ||||||:::|||                  :::||||||        ||
     664 LysAspLysGlyTyrGlyGlySerAlaAspSerGlyProProProLysAs 680

    2376 TCTCATAATAGGC 2388
         |   :::   |||
     680 nAsnMetGluGly 684
```

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT:AAY28655

seq_documentation_block:

ID AAY28655 standard; Protein; 722 AA.
XX
AC AAY28655;
XX
DT 01-OCT-1999 (first entry)
XX
DE Human SVPH1-8 protease.
XX
KW SVPH1-8 protease; testis; fertilisation; spermatogenesis; heredity;
KW adamalysins; ADAMS; snake venom protease; membrane protein; TACE;
KW molecular weight marker; inhibitor; testicular cancer; contraceptive;
KW genetic heterogeneity; genetic disorder; evolutionary tree; reagent;
KW therapeutic agent; detergent; marker; disintegrin domain; antibody.
XX
OS Homo sapiens.

	Key	Location/Qualifiers
	Peptide	1..26
		/label= Signal_sequence
FT	Domain	27..198
FT		/label= Pro_domain
FT	Domain	199..397
FT		/label= Catalytic_domain
FT		/note= "Includes the three conserved His residues"
FT	Domain	398..501
FT		/label= Disintergrin_domain
FT		/note= "Affects fertilisation"
FT	Domain	502..680
FT		/label= Cys_rich_domain
FT	Domain	681..707
FT		/label= Transmembrane_domain
FT	Domain	708..722
FT		/label= Cytoplasmic_domain

XX
PN WO9936549-A1.
XX
PD 22-JUL-1999.
XX
PF 12-JAN-1999; 99WO-US00603.
XX
PR 14-JAN-1998; 98US-0071505.
XX
PA (IMMV) IMMUNEX CORP.
XX
PI Cerretti DP;
XX
DR WPI; 1999-444405/37.
DR N-PSDB; AAX90699.
XX
PT New nucleic acid molecule is useful for screening testicular cells,
PT especially for detection and prognosis of testicular cancer
XX
PS Claim 1b; Page 46-50; 55pp; English.
XX
CC The present protein sequence is the human SVPH1-8 protease, that is
CC specifically expressed in the testis. It may be the human equivalent of
CC adamalysins (ADAMS) essential for fertilisation and spermatogenesis.
CC The catalytic domain is required for the biological activity. The
CC disintegrin domain have been implicated in sperm or egg fusion. It is a
CC member of the snake venom protease family and homologous to the TACE
CC protein, required for shedding of membrane proteins. This sequence is
CC useful for detecting both single and double-stranded RNA and DNA
CC variants of SVPH1-8. Probes specific for SVPH1-8 protease can be used as
CC markers for testis cells and tissues for detection and prognosis of
CC testicular cancer. The fragmented polypeptides are used as markers to
CC determine the molecular weight and isoelectric point of a sample
CC protein. Inhibitors against the disintegrin and catalytic domain can
CC be used as contraceptive agents. The polypeptides are useful for
CC examining genetic heterogeneity and heredity for the identification of
CC risks associated with genetic disorders, to establish evolutionary
CC trees, as an identification reagent, detergent additive and therapeutic
CC agent.
XX
SQ Sequence 722 AA;

alignment_scores:

Quality:	760.00	Length:	721
Ratio:	1.836	Gaps:	32
Percent Similarity:	57.420	Percent Identity:	29.126

alignment_block:
US-09-634-252A-2 x AAY28655 ..

Align seg 1/1 to: AAY28655 from: 1 to: 722

```
     364 AACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACA 413
         :::|||         ||||||     |||:::      |||:::
      34 SerGlnHisPheThrSerProGluValValIleProLeuLysValIleSe 50

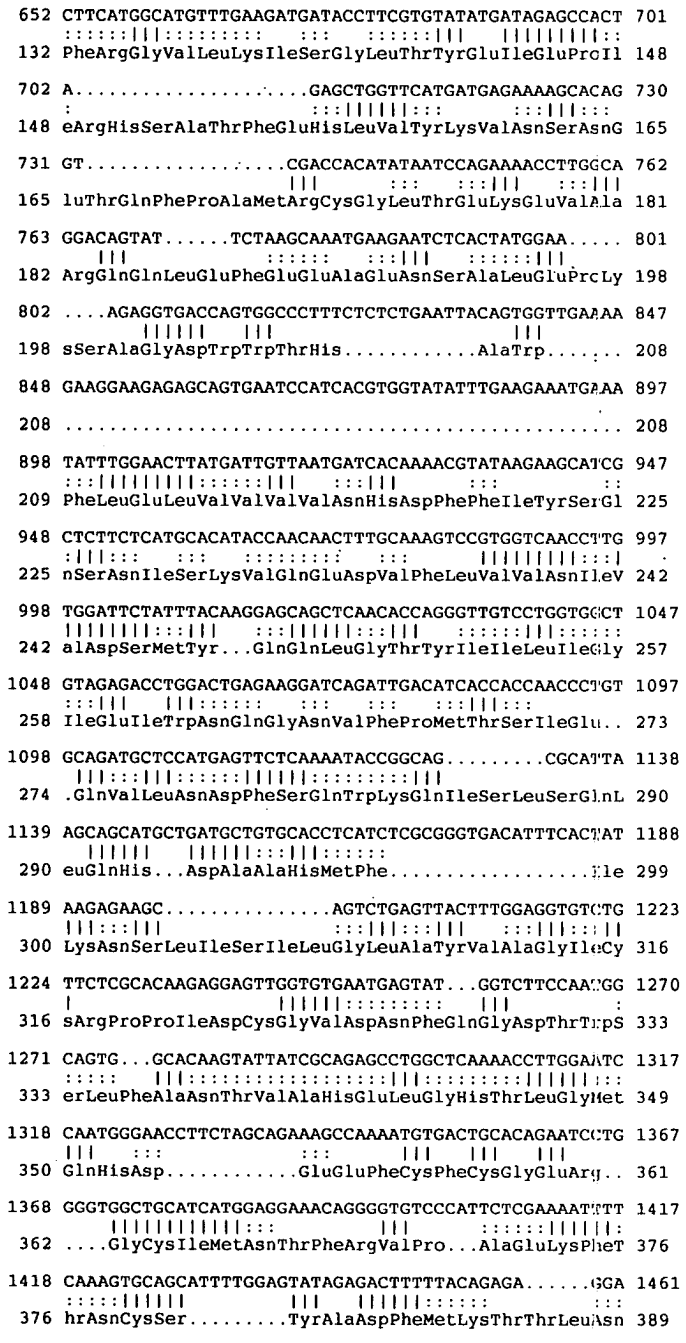
     414 CCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGA 463
          :::   :::      ||||||         :::      |||:::::::
      50 rArgGlyArgSerAlaLysAlaPro......GlyTrpLeuSerTyrSerL 65

     464 TTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGT 513
         :::::      |||:::|||   ::::::   :::   :::::::::
      65 euArgPheGlyGlyGlnLysHisValValHisMetArgValLysLysLeu 81

     514 TTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAAATGGGAAACCACA 563
         |||:::|||              :::   |||   :::   :::
      82 LeuValSerArgHisLeuProValPheThrTyrThrAspGluArgAlaLe 98

     564 GTACTCTAAGGGT.........GGAGAGCACTGTTACTACCATGGAAGCA 604
             :::                  :::   ||||||||||||||||   :
      98 uLeuGluAspGlnLeuPheIleProAspAspCysTyrTyrHisGlyTyrV 115

     605 TCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGC...AATGGA 651
         :::::|||:::   :::|||   |||:::::::|||:::|||   :::|||
     115 alGluGlyAlaProGluSerLeuValValPheSerAlaCysPheGlyGly 131
```





```
     673 SerGlyProAlaSerAlaLysArgGlyValPheLeuProLeuIleValIl 689

    2385 AGGCTCCATCGCT 2397
         |   |||:::::::
     689 eProSerLeuSer 693

seq_name: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT:AAW73007

seq_documentation_block:
ID   AAW73007 standard; Protein; 621 AA.
XX
AC   AAW73007;
XX
DT   02-FEB-1999  (first entry)
XX
DE   Cobra venom protease mocarhagin NMM-1.
XX
KW   Mocarhagin; snake venom; Mozambiquan spitting cobra; protease;
KW   inflammation; myocardial infarction; thrombosis; infection;
     metastasis; therapy; NMM-1.

     Naja mossambica mossambica.

     Key             Location/Qualifiers
FT   Peptide         1..23
FT                   /label= Sig_peptide
FT   Protein         24..621
FT                   /label= Mat_protein
XX
PN   WO9846771-A2.
XX
PD   22-OCT-1998.
XX
PF   14-APR-1998;   98WO-US07998.
XX
PR   18-FEB-1998;   98US-0026001.
PR   15-APR-1997;   97US-0843373.
PR   23-JAN-1998;   98US-0012637.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Boodhoo A,  Sako D,  Seehra JS,  Shaw G;
XX
DR   WPI; 1998-568735/48.
DR   N-PSDB; AAV07895.
XX
PT   Isolated mocarhagin cobra venom protease, and nucleic acids encoding
PT   it - used to develop products for treating e.g. myocardial
PT   infarction, thrombosis, bacterial or viral infection, metastatic
PT   conditions or inflammatory disorders
XX
     Claim 21; Page 35-38; 97pp; English.

     This is the amino acid sequence of mocarhagin NMM-1, a highly
     specific metalloproteinase from the venom of the Mozambiquan
     spitting cobra.  The invention provides mocarhagin polypeptides
CC   (see AAW73007-13) and polynucleotides (see AAV07895-901) encoding them,
CC   as well as host cells and methods of producing the (especially
CC   mature) polypeptides.  Mocarhagin proteins are capable of cleaving
CC   anionic polypeptide containing sulphated tyrosine residues,
CC   P-selectin glycoprotein (GP) ligand-1 (PSGL-1) and GP1b-alpha
CC   (claimed).  They also inhibit neutrophil/HL60 binding, inhibit
CC   platelet binding to von Willebrand Factor, require Ca2+ and Zn2+
CC   ions for activity and have activity inhibited by excess EDTA or
CC   high concentrations of DFP (claimed).  They can be used to inhibit
CC   selectin-mediated binding and to treat inflammatory disease
CC   (claimed).  In particular, they can be used to treat e.g. myocardial
CC   infarction, vessel restenosis, thrombosis, bacterial or viral
CC   infection, metastatic conditions, inflammatory disorders such as
CC   arthritis, acute respiratory distress syndrome, asthma, emphysema,
CC   delayed type hypersensitivity reaction, systemic lupus
CC   erythematosus, thermal injury such as burns or frostbite,
CC   autoimmune thyroiditis, experimental allergic encephalomyelitis,
CC   multiple sclerosis, multiple organ injury syndrome secondary to
CC   trauma, diabetes, Reynaud's syndrome, neutrophilic dermatosis
CC   (Sweet's syndrome), inflammatory bowel disease, Grave's disease,
CC   glomerulonephritis, gingivitis, periodontitis, haemolytic uremic
CC   syndrome, ulcerative colitis, Crohn's disease, necrotising
CC   enterocolitis, granulocyte transfusion associated syndrome,
CC   cytokine-induced enterocolitis, granulocyte transfusion associated
CC   syndrome, or cytokine-induced toxicity.  Mocarhagin protein may
CC   also be useful in organ transplantation, both to prepare organs for
CC   transplantation and to quell organ transplant rejection, to treat
CC   haemodialysis and leukophoresis patients, or as an inhibitor of P-
CC   or E-selectin-mediated intercellular adhesion.
XX
SQ   Sequence  621 AA;


alignment_scores:
          Quality:  751.50            Length:    591
            Ratio:  2.065               Gaps:     19
 Percent Similarity:  61.591   Percent Identity: 30.457

alignment_block:
 US-09-634-252A-2 x AAW73007   ..

 Align seg 1/1  to: AAW73007  from: 1  to: 621

     283 CAGAATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAAT 332
         |||:::||||||      ::::::   :::|||:::   :::   |||::
      17 GlnGlySerSerIleIleLeuGluSerGlyAsnValAsnAspTyrGluVa 33

     333 CACACTGCCTTCAAGACTGATATATTACATCAACCAAGACTCGGAAAGCC 382
         ::::   |||:::::::::           :::          ::::::|
      33 lValTyrProGlnLysValProAlaLeuSerLysGlyGlyValGlnAsnP 50

     383 CTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAG 432
         ||                                  |||   ::::::|||
      50 ro..............................GlnProGluThrLys 55

     433 GCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAA 482
                    |||   :::|||:::::::::      |||:::
      56 TyrGluAspThrMetGlnTyrGluPheHisValAsn......GlyGluPr 70

     483 ATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATG 532
         :::|||   |||      |||:::||||||:::|||:::||||||:
      70 oValValLeuHisLeuGluArgAsnLysGlyLeuPheSerGluAspTyrT 87

     533 TGGAGATTCACTACGAAAATGGGAAACCACAGTACTCTAAGGGTGGA... 579
         ::|||   ||||||              :::   :::   ::::::
      87 hrGluThrHisTyrAlaProAspGlyArgGluIleThrThrSerSerPro 103

     580 ......GAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTC 623
               :::||||||||||||||||||   |||::::::      |||||
     104 ValGlnAspHisCysTyrTyrHisGlyTyrIleGlnAsnGluAlaAspSe 120

     624 CAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATG 673
         |:::::::::::|||:::|||:::||||||   |||   |||:::   :
     120 rSerAlaValIleSerAlaCysAspGlyLeuLysGlyHisPheLysHisG 137

     674 ATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAA 723
         ::      :::|||:::||||||||||||||||||
     137 lnGlyGluThrTyrPheIleGluProLeuGluLeu.............. 148

     724 AGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCAGGACAGTATTC 773
         |||   ::::::   |||   |||
     149 SerAspSerGluAlaHisAlaIle.......................... 156

     774 TAAGCAAATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTC 823
                 |||:::   ::::::|||:::   ::::::   |||   :
     157 .......TyrLysAspGluAsnValGluGluGluGluGluIleProLysI 171

     824 TC.........TCTGAATTACAGTGG..............TTGAAAAGA 849
         ::         ::::::      |||               :::::::::
     171 leCysGlyValThrGlnThrThrTrpGluSerAspGluProIleGluLys 187
```

```
 850 AGGAAGAGAGCAGTGAAT...CCATCACGTGGTATATTT...GAAGAAAT 893
     :::::   :::|||  |||::::::      :::  :::
 188 SerSerGlnLeuThrAsnThrProGluGlnAspArgTyrLeuGlnAlaLy 204

 894 GAAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGC 943
     ||||||:::|||:::  :::|||  |||:::      |||:::|||:
 204 sLysTyrIleGluPheTyrValValValAspAsnValMetTyrArgLysT 221

 944 ATCGCTCTTCT...CATGCACATACCAACAACTTTGCAAAGTCCGTGCTC 990
     ::  ::::::  |||:::  |||:::::: . :::  ::::::|||
 221 yrThrGlyLysLeuHisValIleThrArgArg...ValTyrGluMetVal 236

 991 AACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCT 1040
     |||  :::::::::::::|||:::  :::||||||  :::::::::||
 237 AsnAlaLeuAsnThrMetTyrArg...ArgLeuAsnPheHisIleAlaLe 252

1041 GGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCA 1090
     |:::::::::|||  |||::::::  ::::::|||::::::  ::::
 252 uIleGlyLeuGluIleTrpSerAsnGlyAsnGluIleAsnValGlnSerA 269

1091 ACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGC..... 1134
     ::          |||      |||::::::::::|||::::::
 269 spValGlnAlaThrLeuAspLeuPheGlyGluTrpArgGluAsnLysLeu 285

1135 ...ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATT 1181
          :::::::::  |||  ::::::|||::::::  :::  ||
 286 LeuProArgLysArgAsnAspAsnAlaGlnLeuLeuThrSerThrGluPh 302

1182 TCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCA 1231
     |:::         ::::::|||:::|||:::|||::::::||||||
 302 eAsnGlyThrThrThrGlyLeuGlyTyrIleGlySerLeuCysSerProL 319

1232 CAAGAGGAGTTGGTGTG......AATGAGTATGGTCTTCCAATGGCAGTG 1275
       ::::::|||:::|||     ::::::              |||
 319 ysLysSerValAlaValValGlnAspHisSerLysSerThrSerMetVal 335

1276 GCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGA 1325
     |||  :::::::::::::::::::::::::::|||||||||::::::  ::
 336 AlaIleThrMetAlaHisGlnMetGlyHisAsnLeuGlyMetAsnAspAs 352

1326 ACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCT 1375
     :  :::|||                 ||||||  :::  :::  |
 352 pArgAlaSer.................CysThrCysGlySerAsnLysC 363

1376 GCATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGC 1425
     ||||||||:::             :::          :::||||||:::|||
 363 ysIleMetSerThrLysTyrTyrGluSerLeuSerGluPheSerSerCys 379

1426 AGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCT 1475
     |||:::  |||:::|||::::::|||  |||       |||::
 380 SerValGlnGluHisArgGluTyrLeuLeuArgAspArgProGlnCysIl 396

1476 TTTCAACAGGCCAACAAAG......CTATTTGAGCCCACGGAATGTGGAA 1519
     ::::|||:::|||::::::      :::      |||      |||||||
 396 eLeuAsnLysProSerArgLysAlaIleValThrProProValCysGlyA 413

1520 ATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAA 1569
     ||  :::||||||  ||||||||||||||||||||||         :::
 413 snTyrPheValGluArgGlyGluGluCysAspCysGlySerProGluAsp 429

1570 TGCTATGGATTATGCTGT......AAGAAATGTTCCCTCTCCAACGGGGC 1613
     |||  :::  ||||||            |||:::|||::::::  ||
 430 CysGlnAsnThrCysCysAspAlaAlaThrCysLysLeuGlnHisGluAl 446

1614 TCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGC 1663
     |:::|||::::::|||  ||||||:::      :::|||  |||:::
 446 aGlnCysAspSerGlyGluCysCysGlu......LysCysLysPheLysG 461

1664 CACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAA 1713
       |||  |||||||||  |||  |||:::||||||:::  |||
 461 lyAlaGlyAlaGluCysArgAlaAlaLysAsnAspCysAspPheProGlu 477

1714 TATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGA 1763
       |||||||||  |||::::::||||||  :::      ::::::::
 478 LeuCysThrGlyArgSerAlaLysCysProLysAspSerPheGlnArgAs 494

1764 CGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCA 1813
     :|||:::  |||::::::|||||||||  ||||||||||||  |||
 494 nGlyHisProCysGlnAsnAsnGlnGlyTyrCysTyrAsnGlyThrCysP 511

1814 AGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGG 1863
       |||      |||||||||      :::||||||      |||
 511 roThrLeuThrAsnGlnCysAlaThrLeuTrpGlyProGlyAlaLysMet 527

1864 TCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGG 1913
     |||      :::|||:::      ||||||  ::::::  :::  ||
 528 SerProGlyLeuCysPhe...MetLeuAsnTrpAsnAlaArgSerCysGl 543

1914 AAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATG 1963
     |  |||  |||::::::  |||  |||  |||:::      ||||
 543 yLeuCysArgLysGluAsnGlyArgLysIleLeuCysAlaAlaLysAspV 560

1964 TGTTCTGTGGATTCTTACTCTGT 1986
     ||  ||||||  |||:::|||
 560 alLysCysGlyArgLeuPheCys 567
```

Pts. 1→3 have good data 70% identity

OM of: US-09-634-252A-2 to: Issued_Patents_AA:* out_format : pfs

Date: Aug 15, 2001 9:56 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09634252/runat_14082001_142628_9807/app_query.fasta_1.2598
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=n2p.rai
-GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000
-XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
-GAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
-MATRIX=blosum62 -TRANS=human40.cdi -LIST=90 -DOCALIGN=200
-SORT_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=30 -MODE=LOCAL
-OUTFMT=pfs -NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09634252_@CGN1_1_27 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-634-252A-2
Query length: 2499
Database: Issued_Patents_AA:*
Database sequences: 197339
Database length: 20590346
Search time (sec): 62.900000

→ 70% identity a good Siting data Sir 102(e) application

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	!	Documentation
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-243-542-4 +		1978.00	3104.69	2.1e-166	769	!	S
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-477-407-4 +		1978.00	3104.69	2.1e-166	769	!	S
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-484-355-4 +		1978.00	3104.69	2.1e-166	769	!	S
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-243-542-3 +		1867.50	2931.04	1.1e-156	670	!	S
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-477-407-3 +		1867.50	2931.04	1.1e-156	670	!	S
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-484-355-3 +		1867.50	2931.04	1.1e-156	670	!	S
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-243-542-2 +		1297.00	2029.13	2.5e-106	524	!	S
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-477-407-2 +		1297.00	2029.13	2.5e-106	524	!	S
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-484-355-2 +		1297.00	2029.13	2.5e-106	524	!	S
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-243-542-1 +		1293.00	2023.61	5.4e-106	488	!	S
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-477-407-1 +		1293.00	2023.61	5.4e-106	488	!	S
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-484-355-1 +		1293.00	2023.61	5.4e-106	488	!	S
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-836-443-3 +		829.00	1282.76	6.5e-65	751	!	S
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-765-243-8 +		739.00	1140.29	5.7e-57	734	!	S
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-07295-8 +		739.00	1140.29	5.7e-57	734		
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-264-101-2 +		711.00	1097.28	1.6e-54	651	!	S
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-765-243-2 +		711.00	1097.28	1.6e-54	651	!	S
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-07295-2 +		711.00	1097.28	1.6e-54	651		
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-765-243-6 +		691.00	1064.15	9.9e-53	735	!	S
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-07295-6 +		691.00	1064.15	9.9e-53	735		
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-836-442-3 +		577.50	887.96	9.0e-43	529	!	S
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-264-101-4 +		517.00	793.71	1.8e-37	457	!	S
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-765-243-4 +		517.00	793.71	1.8e-37	457	!	S
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-07295-4 +		517.00	793.71	1.8e-37	457		
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-514-014-2 +		397.50	603.87	6.7e-27	470	!	S
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-833-823-2 +		397.50	603.87	6.7e-27	470	!	S
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-813-150-2 +		397.50	603.87	6.7e-27	470	!	S
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-655-345-4 +		353.50	529.79	6.2e-23	681	!	S
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-09-183-275-4 +		353.50	529.79	6.2e-23	681	!	S
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US96-08407-4 +		353.50	529.79	6.2e-23	681		
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-655-345-2 +		353.50	527.82	6.7e-23	807	!	S
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-09-183-275-2 +		353.50	527.82	6.7e-23	807	!	S
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US96-08407-2 +		353.50	527.82	6.7e-23	807		
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-937-931-8 +		332.50	495.38	4.7e-21	749	!	S
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-285-502-8 +		332.50	495.38	4.7e-21	749	!	S
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-920-234-2 +		331.50	493.81	5.7e-21	748	!	S
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-030-335-9 +		330.50	492.23	7.0e-21	748	!	S
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-030-335-4 +		317.00	470.05	1.1e-19	799	!	S
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-937-931-4 +		315.50	468.44	1.5e-19	748	!	S
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-285-502-4 +		315.50	468.44	1.5e-19	748	!	S
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-706-216-6 +		310.00	467.24	3.3e-19	391	!	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-937-931-2 +		297.00	433.25	8.1e-18	1239		
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-285-502-2 +		297.00	433.25	8.1e-18	1239		
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-030-335-2 +		273.50	406.71	6.2e-16	491	!	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-836-443-4 +		256.50	389.32	1.3e-14	215	!	
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-937-931-6 +		250.50	377.38	4.9e-14	265	!	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-09-285-502-6 +		250.50	377.38	4.9e-14	265	!	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-089-998B-2 +		243.50	367.00	2.0e-13	249		
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-457-272-2 +		243.50	367.00	2.0e-13	249	!	
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US94-07595-2 +		243.50	367.00	2.0e-13	24		
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-11684-2 +		211.00	290.21	4.3e-10	21		
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-11684-4 +		210.50	291.68	4.3e-10	18		
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-836-442-4 +		207.50	312.45	2.7e-10	200	!	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-872-855-10 +		206.50	297.35	5.8e-10	642		
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-400-159-8 +		197.00	276.41	5.2e-09	1065		
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-400-159-10 +		196.00	273.51	6.7e-09	1193		
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-611-729A-10 +		196.00	273.51	6.7e-09	1193		
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-611-729A-8 +		196.00	272.91	6.8e-09	1257		
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-872-855-2 +		193.00	275.19	9.4e-09	685	!	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-882-046-4 +		192.50	268.41	1.3e-08	1148		
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-872-855-4 +		185.50	262.71	4.4e-08	720	!	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-882-046-7 +		185.00	258.00	5.8e-08	1010		
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-400-159-6 +		185.00	255.83	6.3e-08	1218		
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-611-729A-6 +		185.00	255.83	6.3e-08	1218		
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-882-046-2 +		185.00	255.83	6.3e-08	1218		
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-872-855-5 +		181.50	256.48	9.9e-08	713	!	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-882-046-6 +		180.50	248.41	1.6e-07	1248		
/cgn2_6/ptodata/1/iaa/backfiles1.pep:5177197-30 +		180.50	247.13	1.7e-07	1394	!	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-185-432-17 +		180.50	240.10	2.3e-07	2556		
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-185-432-16 +		180.00	239.70	2.5e-07	2471		
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-083-590A-19 +		180.00	239.70	2.5e-07	2471		
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-532-384-19 +		180.00	239.70	2.5e-07	2471		
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-872-855-8 +		179.50	253.05	1.5e-07	729	!	
/cgn2_6/ptodata/1/iaa/6B_COMB.pep:US-08-882-046-5 +		179.00	246.30	2.1e-07	1219		
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-083-590A-20 +		178.50	236.93	3.4e-07	2556		
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-532-384-20 +		178.50	236.93	3.4e-07	2556		
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-185-432-19 +		178.00	235.49	3.9e-07	2703		
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-185-432-18 +		176.00	233.11	5.6e-07	2523		
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-07-965-674-2 +		175.50	271.90	1.2e-07	83	!	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-07-965-674-4 +		175.50	271.90	1.2e-07	83	!	
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US93-09523-2 +		175.50	271.90	1.2e-07	83		
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US93-09523-4 +		175.50	271.90	1.2e-07	83		
/cgn2_6/ptodata/1/iaa/backfiles1.pep:5318899-20 +		175.50	271.90	1.2e-07	83	!	Pa
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-07-965-674-3 +		175.50	271.76	1.2e-07	84	!	
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US93-09523-3 +		175.50	271.76	1.2e-07	84		
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-872-855-7 +		172.00	241.29	6.9e-07	721	!	
/cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-872-855-9 +		171.50	240.56	7.6e-07	717	!	
/cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-07-965-674-1 +		169.00	261.73	4.4e-07	82	!	
/cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US93-09523-1 +		169.00	261.73	4.4e-07	82		
/cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-789-078-1 +		166.50	231.82	2.2e-06	769	!	

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-243-542-4

seq_documentation_block:
; Sequence 4, Application US/08243542
; Patent No. 5552526
; GENERAL INFORMATION:
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: : EMI, MITSURU
; TITLE OF INVENTION: MDC PROTEINS AND DNAs
; TITLE OF INVENTION: ENCODING THE SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLYNN, THIEL, BOUTELL & TANIS P.C.
; STREET: 2026 Rambling Road
; CITY: Kalamazoo
; STATE: Michigan
; COUNTRY: USA
; ZIP: 49008-1699
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage
; COMPUTER: IBM PC/XT/AT Compatible
; OPERATING SYSTEM: MS-DOS 5.0

```
;       SOFTWARE:  WordPerfect 5.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/243,542
;       FILING DATE:
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  JP 5-136602
;       FILING DATE:  14 MAY 1993
;       APPLICATION NUMBER:  JP 5-257455
;       FILING DATE:  22 SEPTEMBER 1993
;       APPLICATION NUMBER:  JP 6-49904
;       FILING DATE:  23 FEBRUARY 1994
;       APPLICATION NUMBER:  JP 6-73328
;       FILING DATE:  12 APRIL 1994
;       APPLICATION NUMBER:  JP 6-84470
;       FILING DATE:  22 APRIL 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Terryence F. Chapman
;       REGISTRATION NUMBER:  32 549
;       REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (616) 381-1156
;       TELEFAX:  (616) 381-5465
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  769 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
;     IMMEDIATE SOURCE:
;       LIBRARY:  human fetal brain cDNA library
US-08-243-542-4


alignment_scores:
            Quality: 1978.00              Length:    817
              Ratio:  3.446                Gaps:     17
 Percent Similarity:  70.257   Percent Identity: 48.592

alignment_block:
 US-09-634-252A-2 x US-08-243-542-4    ..

 Align seg 1/1  to: US-08-243-542-4  from: 1  to: 769

     136 CTGCTTCTCGTCCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCC 185
         |||||||||   ||||||      |||   |||:::::::::::::
      12 LeuLeuLeuSerLeuLeuProThrProGlyLeuGlyThrGlnGlyProAl 28

     186 CCGCGCC......TGGGGGGCT............GCTGCGCCCAGCGCTC 217
              |||      ||||||:::             ::::::|||:::||||
      28 aGlyAlaLeuArgTrpGlyGlyLeuProGlnLeuGlyGlyProGlyAlaP 45

     218 CGCATTGGAATGAAACTGCAGAAAAAAATTTGGGAGTCCTGGCAGATGAA 267
         ||
      45 ro.......................................... 45

     268 GACAATACATTGCAACAGAATAGCAGCAGTAATATCAGTTACAGCAATGC 317

      45 ..................................................... 45

     318 AATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACATCAACC 367
                 |||:::|||   |||||||||||:::          ::::
      46 ..........GluValThrGluProSerArgLeuValArgGlu...SerS 58

     368 AAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAG 417
         ::   :::|||         :::   |||||||||:::::::|||:::::
      58 erGlyGlyGluValArgLysGlnGlnLeuAspThrArgValArgGlnGlu 74

     418 CAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGA 467
                 :::      |||||||||||||||:::||||||   |||
      75 ProProGlyGlyProProValHisLeuAlaGlnValSerPheValIlePr 91

     468 AGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGT 517
         ||||||:::|||:::|||   |||||||||   |||||||:::   ||||
      91 oAlaPheAsnSerAsnPheThrLeuAspLeuGluLeuAsnHisHisLeuL 108

     518 TGTCTTCTGATTATGTGGAGATTCACTAC...GAAAATGGGAAACCACAG 564
         ||||||||:::|||||||||   |||:::   ::::::|||       |||
     108 euSerSerGlnTyrValGluArgHisPheSerArgGluGlyThrThrGln 124

     565 TACTCTAAGGGT...GGAGAGCACTGTTACTACCATGGAAGCATCAGAGG 611
         :::|||   |||   |||:::|||||||||||:::|||:::::|||||
     125 HisSerThrGlyAlaGlyAspHisCysTyrTyrGlnGlyLysLeuArgGl 141

     612 CGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCA 661
         |         |||   :::|||||||||||||||:::|||||||||||:
     141 yAsnProHisSerPheAlaAlaLeuSerThrCysGlnGlyLeuHisGlyV 158

     662 TGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTT 711
         ::|||:::|||   :::::::::|||:::::|||||   |||::::::
     158 alPheSerAspGlyAsnLeuThrTyrIleValGluProGlnGluValAla 174

     712 ......CATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAAC 755
                          :::         ||||||:::|||   :::||
     175 GlyProTrpGlyAlaProGlnGlyProLeuProHisLeuIleTyrArgTh 191

     756 CTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAG 805
         |
     191 r..................................................... 191

     806 GTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAA......... 846
                 |||:::|||   :::           :::
     192 ...........ProLeuLeuProAspProLeuGlyCysArgGluProGly 204

     846 ..................................................... 846

     205 CysLeuPheAlaValProAlaGlnSerAlaProProAsnArgProArgLe 221

     847 .AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT.........ATATTTG 886
          ||||||||||||   |||:::      ||||||         :::   :
     221 uArgArgLysArgGlnValArg......ArgGlyHisProThrValHisS 236

     887 AAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTAT 936
         ::|||   ||||||:::||||||:::::::::||||||||:::   :::
     236 erGluThrLysTyrValGluLeuIleValIleAsnAspHisGlnLeuPhe 252

     937 AAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGT 986
         ::::::   |||:::|||   :::   |||:::||||||||||||||||
     253 GluGlnMetArgGlnSerValValLeuThrSerAsnPheAlaLysSerVa 269

     987 GGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTG 1036
         ||||||||||:::|||   |||||||||||||||||||||||||||:::|
     269 lValAsnLeuAlaAspValIleTyrLysGluGlnLeuAsnThrArgIleV 286

    1037 TCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACC 1086
         ||||||||||:::|||||||||::::::   |||:::|||::::::
     286 alLeuValAlaMetGluThrTrpAlaAspGlyAspLysIleGlnValGln 302

    1087 ACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGC.. 1134
            :::   ::::::   |||   ::::::      ||||||::::::
     303 AspAspLeuLeuGluThrLeuAlaArgLeuMetValTyrArgArgGluGl 319

    1135 .ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTC 1183
          :::   :::   ::::::|||:::|||||:::|||      ||||||:
     319 yLeuProGluProSerAsnAlaThrHisLeuPheSerGlyArgThrPheG 336

    1184 ACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACA 1233
         ::         |||:::   :::|||   ||||||:::||||||   :::
     336 lnSerThrSerSerGlyAlaAlaTyrValGlyGlyIleCysSerLeuSer 352

    1234 AGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACAAGT 1283
         :::|||   ||||||||||||||||         |||:::|||   ::
     353 HisGlyGlyGlyValAsnGluTyrGlyAsnMetGlyAlaMetAlaValTh 369
```

```
1284 ATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGG...GAACCTT 1330
     :|||:::|||:::|||:::|||||||||||||:::::::|||   :::
 369 rLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMetMetTrpAsnLysHisA 386

1331 CTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATC 1380
       |||           |||   |||   :::   |||   |||||||||
 386 rgSerSerAlaGlyAspCysLysCysProAspIleTrpLeuGlyCysIle 402

1381 ATGGAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCAT 1430
     ||||||:::||||||           ||||||||||:::|||||||||
 403 MetGluAspThrGlyPheTyrLeuProArgLysPheSerArgCysSerIl 419

1431 TTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCA 1480
     |   ||||||:::::|||||||||:::|||||||||:::||||||||||
 419 eAspGluTyrAsnGlnPheLeuGlnGluGlyGlyGlySerCysLeuPheA 436

 481 ACAGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATACGTG 1530
     ||:::|||   ||||||::::::|||   |||||||||||||||:::|||
 436 snLysProLeuLysLeuLeuAspProProGluCysGlyAsnGlyPheVal 452

1531 GAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTAT..... 1575
     ||||||||||||||||||||||||||          ||||||
 453 GluAlaGlyGluGluCysAspCysGlySerValGlnGluCysSerArgAl 469

1576 ....GGATTATGCTGTAAGAAATGTTCCCTCTCCAACGGGGCTCACTGCA 1621
         |||   ||||||||||||||:::|||::::::   |||   ||||
 469 aGlyGlyAsnCysCysLysLysCysThrLeuThrHisAspAlaMetCysS 486

1622 GCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGG 1671
     ||||||||   ||||||::::::      |||   ::::::||||||||
 486 erAspGlyLeuCysCysArgArg......CysLysTyrGluProArgGly 500

1672 TATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTAC 1721
        :::||||||:::|||||||||||||||||||||:::|||   |||||
 501 ValSerCysArgGluAlaValAsnGluCysAspIleAlaGluThrCysTh 517

1722 TGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATG 1771
     ||||||||||:::|||||||||||||||||||||||   |||||||||
 517 rGlyAspSerSerGlnCysProProAsnLeuHisLysLeuAspGlyTyrT 534

1772 CATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGA 1821
       |||:::::::::|||||||||||||||||:::|||:::||||||||||||
 534 yrCysAspHisGluGlnGlyArgCysTyrGlyGlyArgCysLysThrArg 550

1822 GACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAA 1871
     |||:::||||||||||   :::||||||      ||||||      |||::
 551 AspArgGlnCysGlnValLeuTrpGlyHisAlaAlaAla......AspAr 565

1872 GTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCG 1921
     :|||||||||||||||||||:::|||||||||||||:::|||:::||||
 565 gPheCysTyrGluLysLeuAsnValGluGlyThrGluArgGlySerCysG 582

1922 GGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGT 1971
     ||:::   |||:::   |||:::||||||||||||:::||||||:::|||
 582 lyArgLysGlySerGlyTrpValGlnCysSerLysGlnAspValLeuCys 598

1972 GGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACT 2021
     ||||||||||||||:::|||:::::   ||||||||||:::|||:::||
 599 GlyPheLeuLeuCysValAsnIleSerGlyAlaProArgLeuGlyAspLe 615

2022 TCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTG 2071
     |   |||:::|||      ::::::||||||||||||||:::   :::|
 615 uValGlyAspIleSerSerValThrPheTyrHisGlnGlyLysGluLeuA 632

2072 ACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTAT 2121
     |||||   |||:::||||||   |||   |||   :::|||::::::|||
 632 spCysArgGlyGlyHisValGlnLeuAlaAspGlySerAspLeuSerTyr 648

2122 GTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCCGAA 2171
     ||||||||||||||   |||||||||:::|||:::||||||||||:::::
 649 ValGluAspGlyThrAlaCysGlyProAsnMetLeuCysLeuAspHisAr 665

2172 GTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATCCA 2221
```

```
     :||||||      :::|||:::|||:::|||:::||||||   :::::::
 665 gCysLeuProAlaSerAlaPheAsnPheSerThrCysProGlySerGlyG 682

2222 AGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGC 2271
     ::   ::::::||||||   |||||||||||||||||||||:::   |||
 682 luArgArgIleCysSerHisHisGlyValCysSerAsnGluGlyLysCys 698

2272 ATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGT 2321
     ||||||:::      |||:::|||   ||||||||||||::::::|||::
 699 IleCysGlnProAspTrpThrGlyLysAspCysSerIleHisAsnProLe 715

2322 TAGGAACCTTCACCCCCCC...AAGGATGAAGGACCCAAGGGTCCTAGTG 2368
     :   :::      |||      :::   |||      ||||||||||||:
 715 uProThrSerProProThrGlyGluThrGluArgTyrLysGlyProSerG 732

2369 CCACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCT 2418
     ::|||||:::|||||||||||||||||||||||||:::||||||||||||
 732 lyThrAsnIleIleIleGlySerIleAlaGlyAlaValLeuValAlaAla 748

2419 ATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAAATGTCAAGAAGAGAAG 2468
     ||||||||||||||||||||||||||||||||||:::::::::   ||
 749 IleValLeuGlyGlyThrGlyTrpGlyPheLysAsnIleArgArgGlyAr 765

2469 G 2469
     |
 765 g 765
```

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-477-407-4

seq_documentation_block:

```
; Sequence 4, Application US/08477407
; Patent No. 5631351
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  :  EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,407
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/243,542
;      FILING DATE:  13-MAY-1994
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
```

```
;       TELEFAX:  (616) 381-5465
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  769 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
;     IMMEDIATE SOURCE:
;       LIBRARY:  human fetal brain cDNA library
US-08-477-407-4


alignment_scores:
           Quality: 1978.00              Length:    817
             Ratio:  3.446                Gaps:     17
 Percent Similarity:  70.257   Percent Identity: 48.592

alignment_block:
[redacted]09-634-252A-2 x US-08-477-407-4   ..

[redacted]n seg 1/1  to: US-08-477-407-4  from: 1  to: 769

     136 CTGCTTCTCGTCCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCC 185
         |||||||||   ||||||      |||   |||:::::::::::
      12 LeuLeuLeuSerLeuLeuProThrProGlyLeuGlyThrGlnGlyProAl 28

     186 CCGCGCC......TGGGGGGCT............GCTGCGCCCAGCGCTC 217
             |||      ||||||:::             ::::::|||:::||||
      28 aGlyAlaLeuArgTrpGlyGlyLeuProGlnLeuGlyGlyProGlyAlaP 45

     218 CGCATTGGAATGAAACTGCAGAAAAAAATTTGGGAGTCCTGGCAGATGAA 267
         ||
      45 ro.......................................... 45

     268 GACAATACATTGCAACAGAATAGCAGCAGTAATATCAGTTACAGCAATGC 317

      45 ..................................................... 45

     318 AATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACATCAACC 367
                 |||:::|||   |||||||||||||:::          ::::
      46 ..........GluValThrGluProSerArgLeuValArgGlu...SerS 58

     368 AAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAG 417
         ::   :::|||         :::   |||||||||::::::|||::::::
      58 erGlyGlyGluValArgLysGlnGlnLeuAspThrArgValArgGlnGlu 74

     418 CAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGA 467
                 :::      |||||||||||||||:::||||||   |||
      75 ProProGlyGlyProProValHisLeuAlaGlnValSerPheValIlePr 91

      68 AGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGT 517
         ||||||:::|||:::|||   |||||||||   ||||||:::   ||||
      91 oAlaPheAsnSerAsnPheThrLeuAspLeuGluLeuAsnHisHisLeuL 108

     518 TGTCTTCTGATTATGTGGAGATTCACTAC...GAAAATGGGAAACCACAG 564
         ||||||||:::|||||||||   |||:::   ::::::|||      |||
     108 euSerSerGlnTyrValGluArgHisPheSerArgGluGlyThrThrGln 124

     565 TACTCTAAGGGT...GGAGAGCACTGTTACTACCATGGAAGCATCAGAGG 611
         :::|||   |||   |||:::||||||||||||:::|||:::::|||||
     125 HisSerThrGlyAlaGlyAspHisCysTyrTyrGlnGlyLysLeuArgGl 141

     612 CGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCA 661
         |         |||   :::|||||||||||||||||:::|||||||||||:
     141 yAsnProHisSerPheAlaAlaLeuSerThrCysGlnGlyLeuHisGlyV 158

     662 TGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTT 711
         ::|||:::|||   :::::::::|||::::::|||||   |||::::::
     158 alPheSerAspGlyAsnLeuThrTyrIleValGluProGlnGluValAla 174

     712 ......CATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAAC 755
                        :::         ||||||:::|||   :::||
     175 GlyProTrpGlyAlaProGlnGlyProLeuProHisLeuIleTyrArgTh 191

     756 CTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAG 805
         |
     191 r..................................................... 191

     806 GTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAA..... 846
                 |||:::|||   :::            :::
     192 ...........ProLeuLeuProAspProLeuGlyCysArgGluProGly 204

     846 ..................................................... 846

     205 CysLeuPheAlaValProAlaGlnSerAlaProProAsnArgProArgLe 221

     847 .AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT.........ATATTTG 886
          ||||||||||||   |||:::      ||||||          :::   :
     221 uArgArgLysArgGlnValArg......ArgGlyHisProThrValHisS 236

     887 AAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTAT 936
         ::|||   ||||||:::|||||::::::::|||||||:::   :::
     236 erGluThrLysTyrValGluLeuIleValIleAsnAspHisGlnLeuPhe 252

     937 AAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGT 986
         ::::::   |||:::|||   :::   |||:::|||||||||||||||||
     253 GluGlnMetArgGlnSerValValLeuThrSerAsnPheAlaLysSerVa 269

     987 GGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTG 1036
         ||||||||||:::|||   |||||||||||||||||||||||||||:::|
     269 lValAsnLeuAlaAspValIleTyrLysGluGlnLeuAsnThrArgIleV 286

    1037 TCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACC 1086
         ||||||||||:::|||||||||::::::   |||:::|||::::::
     286 alLeuValAlaMetGluThrTrpAlaAspGlyAspLysIleGlnValGln 302

    1087 ACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGC.. 1134
           :::   ::::::   |||   ::::::      ||||||::::::
     303 AspAspLeuLeuGluThrLeuAlaArgLeuMetValTyrArgArgGluGl 319

    1135 .ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTC 1183
          :::   :::   :::::::|||:::|||||:::|||      ||||||:
     319 yLeuProGluProSerAsnAlaThrHisLeuPheSerGlyArgThrPheG 336

    1184 ACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACA 1233
         ::         |||:::   :::|||   ||||||:::||||||   :::
     336 lnSerThrSerSerGlyAlaAlaTyrValGlyGlyIleCysSerLeuSer 352

    1234 AGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACAAGT 1283
         :::|||   ||||||||||||||||||          |||:::|||   ::
     353 HisGlyGlyGlyValAsnGluTyrGlyAsnMetGlyAlaMetAlaValTh 369

    1284 ATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGG...GAACCTT 1330
         :|||:::|||:::|||:::||||||||||||:::::::|||   :::
     369 rLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMetMetTrpAsnLysHisA 386

    1331 CTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATC 1380
           |||             |||   |||   :::   |||   ||||||||
     386 rgSerSerAlaGlyAspCysLysCysProAspIleTrpLeuGlyCysIle 402

    1381 ATGGAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCAT 1430
         |||||||:::||||||            ||||||||||||:::||||||||
     403 MetGluAspThrGlyPheTyrLeuProArgLysPheSerArgCysSerIl 419

    1431 TTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCA 1480
         |   ||||||:::::|||||||||:::|||||||||:::||||||||||
     419 eAspGluTyrAsnGlnPheLeuGlnGluGlyGlyGlySerCysLeuPheA 436

    1481 ACAGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATACGTG 1530
         ||:::|||   ||||||:::::|||   |||||||||||||||:::|||
     436 snLysProLeuLysLeuLeuAspProProGluCysGlyAsnGlyPheVal 452

    1531 GAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTAT..... 1575
         |||||||||||||||||||||||||||          ||||||
```

```
 453 GluAlaGlyGluGluCysAspCysGlySerValGlnGluCysSerArgAl 469

1576 ....GGATTATGCTGTAAGAAATGTTCCCTCTCCAACGGGGCTCACTGCA 1621
         |||   ||||||||||||||||:::|||:::::   |||   ||||
 469 aGlyGlyAsnCysCysLysLysCysThrLeuThrHisAspAlaMetCysS 486

1622 GCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGG 1671
     ||||||||   ||||||::::::      |||   ::::::|||||||||
 486 erAspGlyLeuCysCysArgArg......CysLysTyrGluProArgGly 500

1672 TATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTTAC 1721
        :::||||||:::||||||||||||||||||||||:::|||   |||||
 501 ValSerCysArgGluAlaValAsnGluCysAspIleAlaGluThrCysTh 517

1722 TGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATTG 1771
     ||||||||||:::||||||||||||||||||||||   |||||||||
  517 rGlyAspSerSerGlnCysProProAsnLeuHisLysLeuAspGlyTyrT 534

1772 CATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGA 1821
        |||:::::::::|||||||||||||||||:::|||:::||||||||||
  534 yrCysAspHisGluGlnGlyArgCysTyrGlyGlyArgCysLysThrArg 550

1822 GACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAA 1871
     |||:::|||||||||   :::||||||      ||||||      || ::
  551 AspArgGlnCysGlnValLeuTrpGlyHisAlaAlaAla......AspAr 565

1872 GTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCG 1921
     :|||||||||||||||||||:::|||||||||||||:::|||:::|.||
  565 gPheCysTyrGluLysLeuAsnValGluGlyThrGluArgGlySerCysG 582

1922 GGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGGT 1971
     ||:::   |||:::   |||:::|||||||||||||:::|||||||:::|||
  582 lyArgLysGlySerGlyTrpValGlnCysSerLysGlnAspValLeuCys 598

1972 GGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACT 2021
     ||||||||||||||:::|||:::::   |||||||||:::|||:::||
  599 GlyPheLeuLeuCysValAsnIleSerGlyAlaProArgLeuGlyAspLe 615

2022 TCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTG 2071
     |   |||:::|||      :::::::||||||||||||||:::   :::|
  615 uValGlyAspIleSerSerValThrPheTyrHisGlnGlyLysGluLeuA 632

2072 ACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTAT 2121
     |||||   |||:::||||||   |||   |||   :::|||:::::::|||
  632 spCysArgGlyGlyHisValGlnLeuAlaAspGlySerAspLeuSerTyr 648

  122 GTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAA 2171
      ||||||||||||||   |||||||||:::|||:::||||||||:::::
  649 ValGluAspGlyThrAlaCysGlyProAsnMetLeuCysLeuAspHisAr 665

2172 GTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCCA 2221
     :||||||      :::|||:::|||:::|||:::||||||   :::::::
  665 gCysLeuProAlaSerAlaPheAsnPheSerThrCysProGlySerGlyG 682

2222 AGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGC 2271
     ::   ::::::||||||   ||||||||||||||||||||:::   |||
  682 luArgArgIleCysSerHisHisGlyValCysSerAsnGluGlyLysCys 698

2272 ATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGT 2321
     ||||||:::      |||:::|||   ||||||||||||:::::|||::
  699 IleCysGlnProAspTrpThrGlyLysAspCysSerIleHisAsnProLe 715

2322 TAGGAACCTTCACCCCCCC...AAGGATGAAGGACCCAAGGGTCCTAGTG 2368
     :   :::      |||      :::   |||      ||||||||||||:
  715 uProThrSerProProThrGlyGluThrGluArgTyrLysGlyProSerG 732

2369 CCACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCT 2418
     ::||||||:::|||||||||||||||||||||||||:::||||||||||||
  732 lyThrAsnIleIleIleGlySerIleAlaGlyAlaValLeuValAlaAla 748

2419 ATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAAATGTCAAGAAGAGAAG 2468
     ||||||||||||||||||||||||||||||||||:::::::::   ||
  749 IleValLeuGlyGlyThrGlyTrpGlyPheLysAsnIleArgArgGlyAr 765
```

```
    2469 G 2469
         |
     765 g 765
```

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-484-355-4

```
seq_documentation_block:
; Sequence 4, Application US/08484355
; Patent No. 5705341
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  : EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,355
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/243,542
;      FILING DATE:  13-MAY-1994
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
;      TELEFAX:  (616) 381-5465
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  769 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY:  human fetal brain cDNA library
US-08-484-355-4


alignment_scores:
            Quality: 1978.00              Length:    817
              Ratio:  3.446                Gaps:     17
 Percent Similarity:  70.257   Percent Identity: 48.592

alignment_block:
 US-09-634-252A-2 x US-08-484-355-4   ..
```

Align seg 1/1 to: US-08-484-355-4 from: 1 to: 769

```
136 CTGCTTCTCGTCCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCC 185
    |||||||||   ||||||      |||   |||:::::::::::::
 12 LeuLeuLeuSerLeuLeuProThrProGlyLeuGlyThrGlnGlyProAl 28

186 CCGCGCC......TGGGGGGCT............GCTGCGCCCAGCGCTC 217
        |||      ||||||:::             ::::::|||:::|||| 
 28 aGlyAlaLeuArgTrpGlyGlyLeuProGlnLeuGlyGlyProGlyAlaP 45

218 CGCATTGGAATGAAACTGCAGAAAAAAATTTGGGAGTCCTGGCAGATGAA 267
    ||
 45 ro..................................................  45

268 GACAATACATTGCAACAGAATAGCAGCAGTAATATCAGTTACAGCAATGC 317

 45 ..................................................... 45

318 AATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACATCAACC 367
              |||:::|||   |||||||||||:::          ::::
 46 ..........GluValThrGluProSerArgLeuValArgGlu...SerS 58

368 AAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAG 417
    ::   :::|||          :::   |||||||||:::::::|||:::::
 58 erGlyGlyGluValArgLysGlnGlnLeuAspThrArgValArgGlnGlu 74

418 CAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGA 467
              :::      ||||||||||||||||:::|||||   |||
 75 ProProGlyGlyProProValHisLeuAlaGlnValSerPheValIlePr 91

468 AGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGT 517
     ||||||:::|||:::|||   |||||||||   ||||||:::   ||||
 91 oAlaPheAsnSerAsnPheThrLeuAspLeuGluLeuAsnHisHisLeuL 108

518 TGTCTTCTGATTATGTGGAGATTCACTAC...GAAAATGGGAAACCACAG 564
    ||||||||:::|||||||||   |||:::   ::::::|||       |||
108 euSerSerGlnTyrValGluArgHisPheSerArgGluGlyThrThrGln 124

565 TACTCTAAGGGT...GGAGAGCACTGTTACTACCATGGAAGCATCAGAGG 611
    :::|||   |||   |||:::||||||||||||:::|||:::::|||||
125 HisSerThrGlyAlaGlyAspHisCysTyrTyrGlnGlyLysLeuArgGl 141

612 CGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCA 661
    |         |||   :::||||||||||||||:::||||||||||||:
141 yAsnProHisSerPheAlaAlaLeuSerThrCysGlnGlyLeuHisGlyV 158

662 TGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTT 711
    ::|||:::|||   ::::::::::|||:::::|||||   |||::::::
158 alPheSerAspGlyAsnLeuThrTyrIleValGluProGlnGluValAla 174

712 ......CATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAAC 755
                  :::          ||||||:::|||   :::||
175 GlyProTrpGlyAlaProGlnGlyProLeuProHisLeuIleTyrArgTh 191

756 CTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAG 805
    |
191 r..................................................  191

806 GTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAA......... 846
              |||:::|||   :::             :::
192 ..........ProLeuLeuProAspProLeuGlyCysArgGluProGly 204

846 ..................................................  846

205 CysLeuPheAlaValProAlaGlnSerAlaProProAsnArgProArgLe 221

847 .AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT.........ATATTTG 886
     ||||||||||||   |||:::      ||||||         :::   :
221 uArgArgLysArgGlnValArg......ArgGlyHisProThrValHisS 236

887 AAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTAT 936
    ::|||   ||||||:::||||||:::::::::||||||||:::   :::
236 erGluThrLysTyrValGluLeuIleValIleAsnAspHisGlnLeuPhe 252

937 AAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGT 986
    ::::::   |||:::|||   :::   |||:::|||||||||||||||||
253 GluGlnMetArgGlnSerValValLeuThrSerAsnPheAlaLysSerVa 269

987 GGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTG 1036
    ||||||||||:::|||   |||||||||||||||||||||||||||:::|
269 lValAsnLeuAlaAspValIleTyrLysGluGlnLeuAsnThrArgIleV 286

1037 TCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACC 1086
     ||||||||||:::|||||||||::::::   |||:::|||::::::
286 alLeuValAlaMetGluThrTrpAlaAspGlyAspLysIleGlnValGln 302

1087 ACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGC.. 1134
       :::   ::::::   |||   ::::::      ||||||::::::
303 AspAspLeuLeuGluThrLeuAlaArgLeuMetValTyrArgArgGluGl 319

1135 .ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTC 1183
      :::   :::   ::::::|||:::||||||:::|||      ||||||:
319 yLeuProGluProSerAsnAlaThrHisLeuPheSerGlyArgThrPheG 336

1184 ACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACA 1233
     ::         |||:::   :::|||   ||||||:::||||||   :::
336 lnSerThrSerSerGlyAlaAlaTyrValGlyGlyIleCysSerLeuSer 352

1234 AGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACAAGT 1283
     :::|||   ||||||||||||||||||          |||:::|||   ::
353 HisGlyGlyGlyValAsnGluTyrGlyAsnMetGlyAlaMetAlaValTh 369

1284 ATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGG...GAACCTT 1330
     :|||:::|||:::|||:::|||||||||||||:::::::|||   :::
369 rLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMetMetTrpAsnLysHisA 386

1331 CTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATC 1380
       |||            |||   |||   :::   |||   |||||||||
386 rgSerSerAlaGlyAspCysLysCysProAspIleTrpLeuGlyCysIle 402

1381 ATGGAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCAT 1430
     ||||||:::||||||            ||||||||||||:::||||||||
403 MetGluAspThrGlyPheTyrLeuProArgLysPheSerArgCysSerIl 419

1431 TTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCA 1480
     |   ||||||::::::||||||||||:::||||||||||:::||||||||||
419 eAspGluTyrAsnGlnPheLeuGlnGluGlyGlyGlySerCysLeuPheA 436

1481 ACAGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATACGTG 1530
     ||:::|||   ||||||:::::|||   |||||||||||||||:::|||
436 snLysProLeuLysLeuLeuAspProProGluCysGlyAsnGlyPheVal 452

1531 GAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTAT..... 1575
     ||||||||||||||||||||||||||||||          ||||||
453 GluAlaGlyGluGluCysAspCysGlySerValGlnGluCysSerArgAl 469

1576 ....GGATTATGCTGTAAGAAATGTTCCCTCTCCAACGGGGCTCACTGCA 1621
         |||   |||||||||||||||:::|||:::::::   |||   ||||
469 aGlyGlyAsnCysCysLysLysCysThrLeuThrHisAspAlaMetCysS 486

1622 GCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGG 1671
     ||||||||||||||||||:::::      |||   ::::::||||||||
486 erAspGlyLeuCysCysArgArg......CysLysTyrGluProArgGly 500

1672 TATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTAC 1721
        :::||||||:::||||||||||||||||||||||:::|||   |||||
501 ValSerCysArgGluAlaValAsnGluCysAspIleAlaGluThrCysTh 517

1722 TGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATG 1771
     ||||||||||:::|||||||||||||||||||||||   |||||||||
517 rGlyAspSerSerGlnCysProProAsnLeuHisLysLeuAspGlyTyrT 534

1772 CATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGA 1821
       |||:::::::::|||||||||||||:::|||:::||||||||||||
534 yrCysAspHisGluGlnGlyArgCysTyrGlyGlyArgCysLysThrArg 550
```

```
 1822 GACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAA 1871
      |||:::|||||||||   :::||||||      ||||||      |||::
  551 AspArgGlnCysGlnValLeuTrpGlyHisAlaAlaAla.....AspAr 565

 1872 GTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCG 1921
      :|||||||||||||||||||||:::|||||||||||||:::|||:::||||
  565 gPheCysTyrGluLysLeuAsnValGluGlyThrGluArgGlySerCysG 582

 1922 GGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGT 1971
      ||:::  |||:::  |||:::||||||||||||:::||||||:::|||
  582 lyArgLysGlySerGlyTrpValGlnCysSerLysGlnAspValLeuCys 598

 1972 GGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACT 2021
      ||||||||||||||:::|||::::::  |||||||||:::|||:::||
  599 GlyPheLeuLeuCysValAsnIleSerGlyAlaProArgLeuGlyAspLe 615

  022 TCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTG 2071
      |  |||:::|||     ::::::|||||||||||||||:::   :::|
  615 uValGlyAspIleSerSerValThrPheTyrHisGlnGlyLysGluLeuA 632

 2072 ACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTAT 2121
      |||||  |||:::||||||  |||  |||  :::|||:::::|||
  632 spCysArgGlyGlyHisValGlnLeuAlaAspGlySerAspLeuSerTyr 648

 2122 GTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAA 2171
      ||||||||||||||  |||||||||:::|||:::||||||||:::::
  649 ValGluAspGlyThrAlaCysGlyProAsnMetLeuCysLeuAspHisAr 665

 2172 GTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCCA 2221
      :||||||      :::|||:::|||:::|||:::||||||   :::::::
  665 gCysLeuProAlaSerAlaPheAsnPheSerThrCysProGlySerGlyG 682

 2222 AGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGC 2271
      ::   ::::::||||||  ||||||||||||||||||||:::   |||
  682 luArgArgIleCysSerHisHisGlyValCysSerAsnGluGlyLysCys 698

 2272 ATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGT 2321
      ||||||:::      |||:::|||   ||||||||||:::::::||::
  699 IleCysGlnProAspTrpThrGlyLysAspCysSerIleHisAsnProLe 715

 2322 TAGGAACCTTCACCCCCCC...AAGGATGAAGGACCCAAGGGTCCTAGTG 2368
      :   :::      |||      :::   |||      ||||||||||:
  715 uProThrSerProProThrGlyGluThrGluArgTyrLysGlyProSerG 732

 2369 CCACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCT 2418
      ::||||||:::||||||||||||||||||||||:::||||||||||||
  732 lyThrAsnIleIleIleGlySerIleAlaGlyAlaValLeuValAlaAla 748

 2419 ATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAAATGTCAAGAAGAGAAG 2468
      ||||||||||||||||||||||||||||||||||:::::::::   ||
  749 IleValLeuGlyGlyThrGlyTrpGlyPheLysAsnIleArgArgGlyAr 765

 2469 G 2469
      |
  765 g 765
```

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-243-542-3

```
seq_documentation_block:
; Sequence 3, Application US/08243542
; Patent No. 5552526
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/243,542
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
;      TELEFAX:  (616) 381-5465
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  670 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY:  human fetal brain cDNA library
US-08-243-542-3


alignment_scores:
            Quality: 1867.50              Length:    689
              Ratio:  3.640                Gaps:     13
 Percent Similarity:  74.456   Percent Identity: 51.814

alignment_block:
 US-09-634-252A-2 x US-08-243-542-3   ..

 Align seg 1/1  to: US-08-243-542-3  from: 1  to: 670

     502 CTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTAC...GA 548
         |||:::::::   ||||||||||||:::|||||||||   |||:::   ::
       4 LeuSerHisGlnLeuLeuSerSerGlnTyrValGluArgHisPheSerAr 20

     549 AAATGGGAAACCACAGTACTCTAAGGGT...GGAGAGCACTGTTACTACC 595
         ::::|||     |||:::|||   |||   |||:::|||||||||||||:
      20 gGluGlyThrThrGlnHisSerThrGlyAlaGlyAspHisCysTyrTyrG 37

     596 ATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGC 645
         ::|||::::::||||||          |||   :::|||||||||||||||
      37 lnGlyLysLeuArgGlyAsnProHisSerPheAlaAlaLeuSerThrCys 53

     646 AATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGA 695
         :::||||||||||||:::|||:::|||   :::::::::|||::::::||
      54 GlnGlyLeuHisGlyValPheSerAspGlyAsnLeuThrTyrIleValGl 70

     696 GCCACTAGAGCTGGTT......CATGATGAGAAAAGCACAGGTCGACCAC 739
         ||||   |||::::::                  :::         ||||
      70 uProGlnGluValAlaGlyProTrpGlyAlaProGlnGlyProLeuProH 87

     740 ATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAAT 789
         ||:::|||   :::|||
```

```
  87 isLeuIleTyrArgThr..............................  92

 790 CTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTG 839
                                 |||:::|||   :::
  93 ..........................ProLeuLeuProAspProLeuGl 100

 840 GTTGAAA.....................................  846
       :::
 100 yCysArgGluProGlyCysLeuPheAlaValProAlaGlnSerAlaProP 117

 847 .................AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT 879
                     ||||||||||||  |||:::      ||||||
 117 roAsnArgProArgLeuArgArgLysArgGlnValArg......ArgGly 131

 880 .........ATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAA 920
             :::   :::|||   |||||||:::|||||||:::::::::||
 132 HisProThrValHisSerGluThrLysTyrValGluLeuIleValIleAs 148

 921 TGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAACA 970
     |||||||:::   :::::::::   |||:::|||   :::   |||:::|
 148 nAspHisGlnLeuPheGluGlnMetArgGlnSerValValLeuThrSerA 165

 971 ACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAG 1020
     ||||||||||||||||||||||||||:::|||   ||||||||||||||
 165 snPheAlaLysSerValValAsnLeuAlaAspValIleTyrLysGluGln 181

1021 CTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGA 1070
     ||||||||||||:::|||||||||||:::|||||||||::::::   ||
 182 LeuAsnThrArgIleValLeuValAlaMetGluThrTrpAlaAspGlyAs 198

1071 TCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAA 1120
     |:::|||::::::      :::   :::::::   |||   ::::::
 198 pLysIleGlnValGlnAspAspLeuLeuGluThrLeuAlaArgLeuMetV 215

1121 AATACCGGCAGCGC...ATTAAGCAGCATGCTGATGCTGTGCACCTCATC 1167
       ||||||::::::   :::   :::   :::.:::|||:::||||||::
 215 alTyrArgArgGluGlyLeuProGluProSerAsnAlaThrHisLeuPhe 231

1168 TCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGG 1217
     |||       ||||||:::         |||:::   :::|||   |||||
 232 SerGlyArgThrPheGlnSerThrSerSerGlyAlaAlaTyrValGlyGl 248

1218 TGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAA 1267
     |:::||||||   ::::::|||   ||||||||||||||||||
 248 yIleCysSerLeuSerHisGlyGlyGlyValAsnGluTyrGlyAsnMetG 265

1268 TGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATC 1317
       |||:::|||   :::|||:::|||:::|||:::|||||||||||||::
 265 lyAlaMetAlaValThrLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMet 281

1318 CAATGG...GAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATC 1364
     :::|||   :::    |||          |||   |||   :::
 282 MetTrpAsnLysHisArgSerSerAlaGlyAspCysLysCysProAspIl 298

1365 CTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAAT 1414
      |||   ||||||||||||||:::||||||           |||||||
 298 eTrpLeuGlyCysIleMetGluAspThrGlyPheTyrLeuProArgLysP 315

1415 TTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGT 1464
     |||||:::|||||||||   ||||||:::::::||||||||:::||||||
 315 heSerArgCysSerIleAspGluTyrAsnGlnPheLeuGlnGluGlyGly 331

1465 GGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATG 1514
     |||:::||||||||||||:::|||   ||||||::::::|||   |||||
 332 GlySerCysLeuPheAsnLysProLeuLysLeuLeuAspProProGluCy 348

1515 TGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATG 1564
     ||||||||||:::||||||||||||||||||||||||||||||||||||
 348 sGlyAsnGlyPheValGluAlaGlyGluGluCysAspCysGlySerValG 365

1565 TGGAATGCTAT.........GGATTATGCTGTAAGAAATGTTCCCTCTCC 1605
       ||||||            |||   ||||||||||||||:::|||:::
 365 lnGluCysSerArgAlaGlyGlyAsnCysCysLysLysCysThrLeuThr 381

1606 AACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCT 1655
     :::   |||   |||||||||||||   ||||||::::::      |||
  382 HisAspAlaMetCysSerAspGlyLeuCysCysArgArg......CysLy 396

1656 TTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATA 1705
      ::::::|||||||||   :::||||||:::||||||||||||||||||
  396 sTyrGluProArgGlyValSerCysArgGluAlaValAsnGluCysAspI 413

1706 TTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCAT 1755
     ||:::|||   |||||||||||||||:::|||||||||||||||||||||
  413 leAlaGluThrCysThrGlyAspSerSerGlnCysProProAsnLeuHis 429

1756 AAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGG 1805
     |||   |||||||||   |||:::::::::|||||||||||||||:::||
  430 LysLeuAspGlyTyrTyrCysAspHisGluGlnGlyArgCysTyrGlyGl 446

1806 CGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGG 1855
     |:::|||||||||||||||:::|||||||||   :::||||||      |
  446 yArgCysLysThrArgAspArgGlnCysGlnValLeuTrpGlyHisAlaA 463

1856 CTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACT 1905
     |||||      |||:::|||||||||||||||||||||:::|||||||||
  463 laAla......AspArgPheCysTyrGluLysLeuAsnValGluGlyThr 477

1906 GAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAA 1955
     |||:::|||:::||||||:::   |||:::   |||:::||||||||||||
  478 GluArgGlySerCysGlyArgLysGlySerGlyTrpValGlnCysSerLy 494

1956 ACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTC 2005
     |:::|||||||:::|||||||||||||||||||:::|||:::::   ||||
  494 sGlnAspValLeuCysGlyPheLeuLeuCysValAsnIleSerGlyAlaP 511

2006 CACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCAT 2055
     |||||:::|||:::|||   |||:::|||      ::::::|||||||||
  511 roArgLeuGlyAspLeuValGlyAspIleSerSerValThrPheTyrHis 527

2056 CAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGA 2105
     ||||||:::   :::||||||   |||:::||||||   |||   |||
  528 GlnGlyLysGluLeuAspCysArgGlyGlyHisValGlnLeuAlaAspGl 544

2106 TACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGA 2155
      :::|||::::::|||||||||||||||||||   |||||||||:::|||:
  544 ySerAspLeuSerTyrValGluAspGlyThrAlaCysGlyProAsnMetL 561

2156 TGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGC 2205
     ::|||||||||:::::::||||||      :::|||:::|||:::|||:::
  561 euCysLeuAspHisArgCysLeuProAlaSerAlaPheAsnPheSerThr 577

2206 TGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAG 2255
     ||||||   :::::::::   ::::::||||||   ||||||||||||||
  578 CysProGlySerGlyGluArgArgIleCysSerHisHisGlyValCysSe 594

2256 TAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCA 2305
     |||||||:::   ||||||||:::      |||:::|||   |||||||
  594 rAsnGluGlyLysCysIleCysGlnProAspTrpThrGlyLysAspCysS 611

2306 GTATCCGGGATCCAGTTAGGAACCTTCACCCCCCC...AAGGATGAAGGA 2352
     |||||:::::|||:::   :::      |||      :::   |||
  611 erIleHisAsnProLeuProThrSerProProThrGlyGluThrGluArg 627

2353 CCCAAGGGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGGTGC 2402
       ||||||||||||:::||||||:::|||||||||||||||||||||||
  628 TyrLysGlyProSerGlyThrAsnIleIleIleGlySerIleAlaGlyAl 644

2403 CATCCTGGTAGCAGCTATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAA 2452
     |:::|||||||||||||||||||||||||||||||||||||||||||||
  644 aValLeuValAlaAlaIleValLeuGlyGlyThrGlyTrpGlyPheLysA 661

2453 ATGTCAAGAAGAGAAGG 2469
     ||:::::::::   |||
  661 snIleArgArgGlyArg 666
```

```
seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-477-407-3

seq_documentation_block:
; Sequence 3, Application US/08477407
; Patent No. 5691351
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  :  EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,407
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/243,542
;      FILING DATE:  13-MAY-1994
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
;      TELEFAX:  (616) 381-5465
;    FORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  670 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY:  human fetal brain cDNA library
US-08-477-407-3


alignment_scores:
            Quality: 1867.50              Length:    689
              Ratio:  3.640               Gaps:       13
 Percent Similarity:  74.456   Percent Identity: 51.814

alignment_block:
 US-09-634-252A-2 x US-08-477-407-3   ..

 Align seg 1/1  to: US-08-477-407-3  from: 1  to: 670

     502 CTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTAC...GA 548
         |||::::::   |||||||||||||:::|||||||||   |||:::   ::
       4 LeuSerHisGlnLeuLeuSerSerGlnTyrValGluArgHisPheSe::Ar 20

     549 AAATGGGAAACCACAGTACTCTAAGGGT...GGAGAGCACTGTTACTACC 595
         ::::|||      |||:::|||   |||   |||:::|||||||||||||:
      20 gGluGlyThrThrGlnHisSerThrGlyAlaGlyAspHisCysTyrTyrG 37

     596 ATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGC 645
         ::|||::::::||||||         |||   :::|||||||||||||||
      37 lnGlyLysLeuArgGlyAsnProHisSerPheAlaAlaLeuSerThrCys 53

     646 AATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGA 695
         :::|||||||||||||:::|||:::|||   :::::::::|||::::::||
      54 GlnGlyLeuHisGlyValPheSerAspGlyAsnLeuThrTyrIleValGl 70

     696 GCCACTAGAGCTGGTT......CATGATGAGAAAAGCACAGGTCGACCAC 739
         ||||   |||::::::                    :::         ||||
      70 uProGlnGluValAlaGlyProTrpGlyAlaProGlnGlyProLeuProH 87

     740 ATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAAT 789
         ||:::|||   :::|||
      87 isLeuIleTyrArgThr..............................  92

     790 CTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTG 839
                                     |||:::|||   :::
      93 ..........................ProLeuLeuProAspProLeuGl 100

     840 GTTGAAA.....................................  846
             :::
     100 yCysArgGluProGlyCysLeuPheAlaValProAlaGlnSerAlaProP 117

     847 ................AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT 879
                         ||||||||||||  |||:::      ||||||
     117 roAsnArgProArgLeuArgArgLysArgGlnValArg......ArgGly 131

     880 .......ATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAA 920
                 :::   :::|||   ||||||:::||||||::::::::||
     132 HisProThrValHisSerGluThrLysTyrValGluLeuIleValIleAs 148

     921 TGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAACA 970
         |||||||:::   :::::::::   |||:::|||   :::   |||:::|
     148 nAspHisGlnLeuPheGluGlnMetArgGlnSerValValLeuThrSerA 165

     971 ACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAG 1020
         ||||||||||||||||||||||||||:::|||   ||||||||||||||||
     165 snPheAlaLysSerValValAsnLeuAlaAspValIleTyrLysGluGln 181

    1021 CTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGA 1070
         ||||||||||||:::||||||||||||:::|||||||||::::::   ||
     182 LeuAsnThrArgIleValLeuValAlaMetGluThrTrpAlaAspGlyAs 198

    1071 TCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAA 1120
         |:::|||::::::      :::   ::::::   |||   ::::::
     198 pLysIleGlnValGlnAspAspLeuLeuGluThrLeuAlaArgLeuMetV 215

    1121 AATACCGGCAGCGC...ATTAAGCAGCATGCTGATGCTGTGCACCTCATC 1167
           ||||||::::::   :::   :::   :::::::|||:::||||||:::
     215 alTyrArgArgGluGlyLeuProGluProSerAsnAlaThrHisLeuPhe 231

    1168 TCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGG 1217
         |||      ||||||:::         |||:::   :::|||   |||||
     232 SerGlyArgThrPheGlnSerThrSerSerGlyAlaAlaTyrValGlyGl 248

    1218 TGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAA 1267
         |:::||||||   :::::::|||   |||||||||||||||||
     248 yIleCysSerLeuSerHisGlyGlyGlyValAsnGluTyrGlyAsnMetG 265

    1268 TGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATC 1317
           |||:::|||   :::|||:::|||:::|||:::|||||||||||||:::
     265 lyAlaMetAlaValThrLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMet 281

    1318 CAATGG...GAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATC 1364
         :::|||   :::      |||           |||   |||   :::
     282 MetTrpAsnLysHisArgSerSerAlaGlyAspCysLysCysProAspIl 298
```

```
1365 CTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAAT 1414
      |||   |||||||||||||||:::||||||           |||||||
 298 eTrpLeuGlyCysIleMetGluAspThrGlyPheTyrLeuProArgLysP 315

1415 TTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGT 1464
     |||||:::|||||||||   ||||||:::::::|||||||||:::||||||
 315 heSerArgCysSerIleAspGluTyrAsnGlnPheLeuGlnGluGlyGly 331

1465 GGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATG 1514
     |||:::||||||||||||:::|||   ||||||:::::|||   |||||
 332 GlySerCysLeuPheAsnLysProLeuLysLeuLeuAspProProGluCy 348

1515 TGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATG 1564
     ||||||||||:::|||||||||||||||||||||||||||||
 348 sGlyAsnGlyPheValGluAlaGlyGluGluCysAspCysGlySerValG 365

1565 TGGAATGCTAT.........GGATTATGCTGTAAGAAATGTTCCCTCTCC 1605
       ||||||            |||   |||||||||||||||:::|||:::
 365 lnGluCysSerArgAlaGlyGlyAsnCysCysLysLysCysThrLeuThr 381

 506 AACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCT 1655
     :::   |||   ||||||||||||   ||||||:::::::      |||
 382 HisAspAlaMetCysSerAspGlyLeuCysCysArgArg......CysLy 396

1656 TTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATA 1705
      ::::::|||||||||   :::|||||||:::||||||||||||||||||
 396 sTyrGluProArgGlyValSerCysArgGluAlaValAsnGluCysAspI 413

1706 TTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCAT 1755
     ||:::|||   |||||||||||||||:::|||||||||||||||||||||
 413 leAlaGluThrCysThrGlyAspSerSerGlnCysProProAsnLeuHis 429

1756 AAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGG 1805
     |||   |||||||||   |||:::::::::|||||||||||||||:::||
 430 LysLeuAspGlyTyrTyrCysAspHisGluGlnGlyArgCysTyrGlyGl 446

1806 CGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGG 1855
     |:::|||||||||||||||:::|||||||||   :::||||||      |
 446 yArgCysLysThrArgAspArgGlnCysGlnValLeuTrpGlyHisAlaA 463

1856 CTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACT 1905
     |||||      |||:::|||||||||||||||||||||:::|||||||||
 463 laAla......AspArgPheCysTyrGluLysLeuAsnValGluGlyThr 477

1906 GAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAA 1955
     |||:::|||:::||||||:::   |||:::   |||:::|||||||||||
 478 GluArgGlySerCysGlyArgLysGlySerGlyTrpValGlnCysSerLy 494

1956 ACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTC 2005
     |:::||||||:::|||||||||||||||||||:::|||:::::   ||||
 494 sGlnAspValLeuCysGlyPheLeuLeuCysValAsnIleSerGlyAlaP 511

2006 CACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCAT 2055
     |||||:::|||:::|||   |||:::|||      ::::::||||||||
2511 roArgLeuGlyAspLeuValGlyAspIleSerSerValThrPheTyrHis 527

2056 CAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGA 2105
     |||||||::   :::||||||   |||:::||||||   |||   |||
 528 GlnGlyLysGluLeuAspCysArgGlyGlyHisValGlnLeuAlaAspGl 544

2106 TACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGA 2155
      :::|||:::::|||||||||||||||||||   ||||||||||:::|||:
 544 ySerAspLeuSerTyrValGluAspGlyThrAlaCysGlyProAsnMetL 561

2156 TGTGTTTAGATGGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGC 2205
     ::|||||||||::::||||||      :::|||:::|||:::|||:::
 561 euCysLeuAspHisArgCysLeuProAlaSerAlaPheAsnPheSerThr 577

2206 TGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAG 2255
     ||||||   :::::::::   ::::::||||||   ||||||||||||||
 578 CysProGlySerGlyGluArgArgIleCysSerHisHisGlyValCysSe 594

2256 TAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCA 2305
```

```
        |||||||:::   ||||||||:::      |||:::|||   |||||||
 594 rAsnGluGlyLysCysIleCysGlnProAspTrpThrGlyLysAspCysS 611

2306 GTATCCGGGATCCAGTTAGGAACCTTCACCCCCCC...AAGGATGAAGGA 2352
     |||||::::::|||:::   :::      |||      :::   |||
 611 erIleHisAsnProLeuProThrSerProProThrGlyGluThrGluArg 627

2353 CCCAAGGGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGGTGC 2402
       ||||||||||||:::|||||||:::|||||||||||||||||||||||
 628 TyrLysGlyProSerGlyThrAsnIleIleIleGlySerIleAlaGlyAl 644

2403 CATCCTGGTAGCAGCTATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAA 2452
     |:::|||||||||||||||||||||||||||||||||||||||||||||
 644 aValLeuValAlaAlaIleValLeuGlyGlyThrGlyTrpGlyPheLysA 661

2453 ATGTCAAGAAGAGAAGG 2469
     ||:::::::::   |||
 661 snIleArgArgGlyArg 666
```

```
seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-484-355-3

seq_documentation_block:
; Sequence 3, Application US/08484355
; Patent No. 5705341
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  : EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,355
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/243,542
;      FILING DATE:  13-MAY-1994
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
;      TELEFAX:  (616) 381-5465
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  670 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
```

```
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY:  human fetal brain cDNA library
US-08-484-355-3


alignment_scores:
            Quality: 1867.50              Length:    689
              Ratio:  3.640                Gaps:     13
 Percent Similarity:  74.456   Percent Identity: 51.814

alignment_block:
 US-09-634-252A-2 x US-08-484-355-3   ..

 Align seg 1/1  to: US-08-484-355-3  from: 1  to: 670

  502 CTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTAC...GA 548
        |||::::::   |||||||||||||:::|||||||||   |||:::   ::
    4 LeuSerHisGlnLeuLeuSerSerGlnTyrValGluArgHisPheSerAr 20

  549 AAATGGGAAACCACAGTACTCTAAGGGT...GGAGAGCACTGTTACTACC 595
        ::::|||      |||:::|||   |||   |||:::|||||||||||:
   20 gGluGlyThrThrGlnHisSerThrGlyAlaGlyAspHisCysTyrTyrG 37

  596 ATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGC 645
        ::|||:::::|||||          |||   :::|||||||||||||||
   37 lnGlyLysLeuArgGlyAsnProHisSerPheAlaAlaLeuSerThrCys 53

  646 AATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGA 695
        :::|||||||||||||:::|||:::|||   :::::::::|||:::::||
   54 GlnGlyLeuHisGlyValPheSerAspGlyAsnLeuThrTyrIleValGl 70

  696 GCCACTAGAGCTGGTT......CATGATGAGAAAAGCACAGGTCGACCAC 739
        ||||   |||::::::              :::          ||||
   70 uProGlnGluValAlaGlyProTrpGlyAlaProGlnGlyProLeuProH 87

  740 ATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAAT 789
        ||:::|||   :::|||
   87 isLeuIleTyrArgThr..................................... 92

  790 CTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTG 839
                                  |||:::|||   :::
   93 ..........................ProLeuLeuProAspProLeuGl 100

  840 GTTGAAA.................................................. 846
          :::
  100 yCysArgGluProGlyCysLeuPheAlaValProAlaGlnSerAlaProP 117

  847 ................AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT 879
                      ||||||||||||   |||:::      ||||||
  117 roAsnArgProArgLeuArgArgLysArgGlnValArg......ArgGly 131

  880 .........ATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAA 920
              :::   :::|||   ||||||:::||||||:::::::::||
  132 HisProThrValHisSerGluThrLysTyrValGluLeuIleValIleAs 148

  921 TGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAACA 970
        |||||||:::   :::::::::   |||:::|||   :::   |||:::|
  148 nAspHisGlnLeuPheGluGlnMetArgGlnSerValValLeuThrSerA 165

  971 ACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAG 1020
        ||||||||||||||||||||||||||:::|||   ||||||||||||||
  165 snPheAlaLysSerValValAsnLeuAlaAspValIleTyrLysGluGln 181

 1021 CTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGA 1070
        ||||||||||||:::||||||||||||:::||||||||:::::::   ||
  182 LeuAsnThrArgIleValLeuValAlaMetGluThrTrpAlaAspGlyAs 198

 1071 TCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAA 1120
        |:::|||:::::       :::   ::::::   |||   ::::::
  198 pLysIleGlnValGlnAspAspLeuLeuGluThrLeuAlaArgLeuMetV 215

 1121 AATACCGGCAGCGC...ATTAAGCAGCATGCTGATGCTGTGCACCTCATC 1167
        ||||||::::::   :::   :::   ::::::|||:::||||||:::
  215 alTyrArgArgGluGlyLeuProGluProSerAsnAlaThrHisLeuPhe 231

 1168 TCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGG 1217
        |||      |||||||:::         |||:::   :::|||   |||||
  232 SerGlyArgThrPheGlnSerThrSerSerGlyAlaAlaTyrValGlyGl 248

 1218 TGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAA 1267
        |:::||||||   ::::::|||   ||||||||||||||||||
  248 yIleCysSerLeuSerHisGlyGlyGlyValAsnGluTyrGlyAsnMetG 265

 1268 TGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATC 1317
        |||:::|||  .:::|||:::|||:::|||:::||||||||||||:::
  265 lyAlaMetAlaValThrLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMet 281

 1318 CAATGG...GAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATC 1364
        :::|||   :::      |||            |||   |||   :::
  282 MetTrpAsnLysHisArgSerSerAlaGlyAspCysLysCysProAspIl 298

 1365 CTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAAT 1414
        |||   |||||||||||||||:::||||||            |||||||
  298 eTrpLeuGlyCysIleMetGluAspThrGlyPheTyrLeuProArgLysP 315

 1415 TTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGT 1464
        |||||:::|||||||||   ||||||:::::||||||||:::||||||
  315 heSerArgCysSerIleAspGluTyrAsnGlnPheLeuGlnGluGlyGly 331

 1465 GGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATG 1514
        |||:::|||||||||||:::|||   ||||||:::::|||   |||||
  332 GlySerCysLeuPheAsnLysProLeuLysLeuLeuAspProProGluCy 348

 1515 TGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATG 1564
        ||||||||||:::||||||||||||||||||||||||||||||||
  348 sGlyAsnGlyPheValGluAlaGlyGluGluCysAspCysGlySerValG 365

 1565 TGGAATGCTAT.........GGATTATGCTGTAAGAAATGTTCCCTCTCC 1605
        ||||||            |||   |||||||||||||||:::|||:::
  365 lnGluCysSerArgAlaGlyGlyAsnCysCysLysLysCysThrLeuThr 381

 1606 AACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCT 1655
        :::  |||   ||||||||||||   ||||||::::::      |||
  382 HisAspAlaMetCysSerAspGlyLeuCysCysArgArg......CysLy 396

 1656 TTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATA 1705
        ::::::|||||||||   :::||||||:::||||||||||||||||||
  396 sTyrGluProArgGlyValSerCysArgGluAlaValAsnGluCysAspI 413

 1706 TTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCAT 1755
        ||:::|||   ||||||||||||||:::||||||||||||||||||||
  413 leAlaGluThrCysThrGlyAspSerSerGlnCysProProAsnLeuHis 429

 1756 AAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGG 1805
        |||   |||||||||   |||:::::::::||||||||||||||:::||
  430 LysLeuAspGlyTyrTyrCysAspHisGluGlnGlyArgCysTyrGlyGl 446

 1806 CGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGG 1855
        |:::|||||||||||||||:::|||||||||   :::||||||      |
  446 yArgCysLysThrArgAspArgGlnCysGlnValLeuTrpGlyHisAlaA 463

 1856 CTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACT 1905
        |||||      |||:::|||||||||||||||||||||:::|||||||||
  463 laAla......AspArgPheCysTyrGluLysLeuAsnValGluGlyThr 477

 1906 GAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAA 1955
        |||:::|||:::|||||||:::   |||:::   |||:::||||||||||
  478 GluArgGlySerCysGlyArgLysGlySerGlyTrpValGlnCysSerLy 494

 1956 ACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTC 2005
        |:::||||||:::|||||||||||||||||||:::|||:::::   ||||
  494 sGlnAspValLeuCysGlyPheLeuLeuCysValAsnIleSerGlyAlaP 511

 2006 CACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCAT 2055
        |||||:::|||:::|||   |||:::|||      ::::::|||||||||
```

```
     511 roArgLeuGlyAspLeuValGlyAspIleSerSerValThrPheTyrHis 527

    2056 CAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGA 2105
         |||||||:::   :::||||||   |||:::||||||   |||   |||
     528 GlnGlyLysGluLeuAspCysArgGlyGlyHisValGlnLeuAlaAspGl 544

    2106 TACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGA 2155
          :::|||:::::|||||||||||||||||||   ||||||||||:::|||:
     544 ySerAspLeuSerTyrValGluAspGlyThrAlaCysGlyProAsnMetL 561

    2156 TGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGC 2205
         ::|||||||||:::::|||||      :::|||:::|||:::|||:::
     561 euCysLeuAspHisArgCysLeuProAlaSerAlaPheAsnPheSerThr 577

    2206 TGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAG 2255
         ||||||   :::::::::   ::::::||||||   ||||||||||||||
     578 CysProGlySerGlyGluArgArgIleCysSerHisHisGlyValCysSe 594

    2256 TAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCA 2305
         |||||||:::   |||||||||:::      |||:::|||   |||||||
     594 rAsnGluGlyLysCysIleCysGlnProAspTrpThrGlyLysAspCysS 611

     306 GTATCCGGGATCCAGTTAGGAACCTTCACCCCCCC...AAGGATGAAGGA 2352
         |||||:::::::|||:::   :::      |||      :::   |||
     611 erIleHisAsnProLeuProThrSerProProThrGlyGluThrGluArg 627

    2353 CCCAAGGGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGGTGC 2402
          ||||||||||||:::||||||:::||||||||||||||||||||||||
     628 TyrLysGlyProSerGlyThrAsnIleIleIleGlySerIleAlaGlyAl 644

    2403 CATCCTGGTAGCAGCTATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAA 2452
         |:::||||||||||||||||||||||||||||||||||||||||||||||
     644 aValLeuValAlaAlaIleValLeuGlyGlyThrGlyTrpGlyPheLysA 661

    2453 ATGTCAAGAAGAGAAGG 2469
         ||:::::::::   |||
     661 snIleArgArgGlyArg 666

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-243-542-2

seq_documentation_block:
; Sequence 2, Application US/08243542
; Patent No. 5552526
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  :  EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
:      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
       STREET:  2026 Rambling Road
       CITY:  Kalamazoo
       STATE:  Michigan
       COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/243,542
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
;      TELEFAX:  (616) 381-5465
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  524 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY:  human fetal brain cDNA library
US-08-243-542-2


alignment_scores:
            Quality: 1297.00              Length:    517
              Ratio:  3.515                Gaps:     12
 Percent Similarity:  71.373   Percent Identity: 49.516

alignment_block:
 US-09-634-252A-2 x US-08-243-542-2   ..

 Align seg 1/1  to: US-08-243-542-2  from: 1  to: 524

     502 CTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTAC...GA 548
         |||::::::   ||||||||||||:::|||||||||   |||:::   ::
       4 LeuSerHisGlnLeuLeuSerSerGlnTyrValGluArgHisPheSerAr 20

     549 AAATGGGAAACCACAGTACTCTAAGGGT...GGAGAGCACTGTTACTACC 595
         ::::|||      |||:::|||   |||   |||:::||||||||||||:
      20 gGluGlyThrThrGlnHisSerThrGlyAlaGlyAspHisCysTyrTyrG 37

     596 ATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGC 645
         ::|||::::::|||||      |||   :::||||||||||||||||
      37 lnGlyLysLeuArgGlyAsnProHisSerPheAlaAlaLeuSerThrCys 53

     646 AATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGA 695
         :::||||||||||||:::|||:::|||   ::::::::::|||:::::||
      54 GlnGlyLeuHisGlyValPheSerAspGlyAsnLeuThrTyrIleValGl 70

     696 GCCACTAGAGCTGGTT......CATGATGAGAAAAGCACAGGTCGACCAC 739
         ||||   |||::::::                  :::         ||||
      70 uProGlnGluValAlaGlyProTrpGlyAlaProGlnGlyProLeuProH 87

     740 ATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAAT 789
         ||:::|||   :::|||
      87 isLeuIleTyrArgThr..............................  92

     790 CTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTG 839
                                      |||:::|||   :::
      93 ..........................ProLeuLeuProAspProLeuGl 100

     840 GTTGAAA.....................................  846
         :::
     100 yCysArgGluProGlyCysLeuPheAlaValProAlaGlnSerAlaProP 117

     847 ................AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT 879
                        ||||||||||||   |||:::      ||||||
     117 roAsnArgProArgLeuArgArgLysArgGlnValArg......ArgGly 131

     880 .........ATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAA 920
                 :::   :::|||   ||||||:::||||||:::::::::||
     132 HisProThrValHisSerGluThrLysTyrValGluLeuIleValIleAs 148
```

```
 921 TGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAACA 970
     |||||||:::   :::::::::   |||:::|||   :::   |||:::|
 148 nAspHisGlnLeuPheGluGlnMetArgGlnSerValValLeuThrSerA 165

 971 ACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAG 1020
     ||||||||||||||||||||||||||:::|||   ||||||||||||||
 165 snPheAlaLysSerValValAsnLeuAlaAspValIleTyrLysGluGln 181

1021 CTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGA 1070
     ||||||||||||:::|||||||||||:::|||||||||::::::   ||
 182 LeuAsnThrArgIleValLeuValAlaMetGluThrTrpAlaAspGlyAs 198

1071 TCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAA 1120
     |:::|||::::::      :::   ::::::   |||   ::::::
 198 pLysIleGlnValGlnAspAspLeuLeuGluThrLeuAlaArgLeuMetV 215

1121 AATACCGGCAGCGC...ATTAAGCAGCATGCTGATGCTGTGCACCTCATC 1167
       ||||||::::::   :::   :::   :::::::|||:::|||||:::
 215 alTyrArgArgGluGlyLeuProGluProSerAsnAlaThrHisLeuPhe 231

1168 TCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGG 1217
     |||      ||||||:::          |||:::   :::|||   |||||
 232 SerGlyArgThrPheGlnSerThrSerSerGlyAlaAlaTyrValGlyGl 248

1218 TGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAA 1267
     |:::||||||   :::::::|||   |||||||||||||||||
 248 yIleCysSerLeuSerHisGlyGlyGlyValAsnGluTyrGlyAsnMetG 265

1268 TGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATC 1317
       |||:::|||   :::|||:::|||:::|||:::|||||||||||:::
 265 lyAlaMetAlaValThrLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMet 281

1318 CAATGG...GAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATC 1364
     :::|||   :::      |||          |||   |||   :::
 282 MetTrpAsnLysHisArgSerSerAlaGlyAspCysLysCysProAspIl 298

1365 CTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAAT 1414
      |||   ||||||||||||||:::|||||||           |||||||
 298 eTrpLeuGlyCysIleMetGluAspThrGlyPheTyrLeuProArgLysP 315

1415 TTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAAGT 1464
     |||||:::|||||||||   ||||||:::::|||||||||:::||||||
 315 heSerArgCysSerIleAspGluTyrAsnGlnPheLeuGlnGluGlyGly 331

1465 GGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATG 1514
     |||:::|||||||||||||:::|||   ||||||::::::|||   |||||
 332 GlySerCysLeuPheAsnLysProLeuLysLeuLeuAspProProGluCy 348

1515 TGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATG 1564
     ||||||||||:::|||||||||||||||||||||||||||||||||||
 348 sGlyAsnGlyPheValGluAlaGlyGluGluCysAspCysGlySerValG 365

1565 TGGAATGCTAT.........GGATTATGCTGTAAGAAATGTTCCCTCTCC 1605
       ||||||            |||   |||||||||||||||:::|||:::
 365 lnGluCysSerArgAlaGlyGlyAsnCysCysLysLysCysThrLeuThr 381

1606 AACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCT 1655
     :::   |||   ||||||||||||   ||||||::::::      |||
 382 HisAspAlaMetCysSerAspGlyLeuCysCysArgArg......CysLy 396

1656 TTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATA 1705
      ::::::|||||||||   :::|||||:::|||||||||||||||||||
 396 sTyrGluProArgGlyValSerCysArgGluAlaValAsnGluCysAspI 413

1706 TTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCAT 1755
     ||:::|||   |||||||||||||||:::|||||||||||||||||||||
 413 leAlaGluThrCysThrGlyAspSerSerGlnCysProProAsnLeuHis 429

1756 AAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGG 1805
     |||   |||||||||   |||:::::::::|||||||||||||||:::||
 430 LysLeuAspGlyTyrTyrCysAspHisGluGlnGlyArgCysTyrGlyGl 446

1806 CGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGG 1855
```

```
          |:::|||||||||||||||:::|||||||||   :::||||||      |
      446 yArgCysLysThrArgAspArgGlnCysGlnValLeuTrpGlyHisAlaA 463

     1856 CTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACT 1905
          |||||      |||:::|||||||||||||||||||||:::|||||||||
      463 laAla......AspArgPheCysTyrGluLysLeuAsnValGluGlyThr 477

     1906 GAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAA 1955
          |||:::|||:::||||||:::   |||:::   |||:::||||||||||
      478 GluArgGlySerCysGlyArgLysGlySerGlyTrpValGlnCysSerLy 494

     1956 A 1956
          |
      494 s 494
```

```
seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-477-407-2

seq_documentation_block:
; Sequence 2, Application US/08477407
; Patent No. 5631351
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  :  EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,407
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/243,542
;      FILING DATE:  13-MAY-1994
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
;      TELEFAX:  (616) 381-5465
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  524 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY:  human fetal brain cDNA library
US-08-477-407-2
```

```
alignment_scores:
            Quality: 1297.00              Length:    517
              Ratio:  3.515                Gaps:     12
 Percent Similarity:  71.373   Percent Identity: 49.516

alignment_block:
 US-09-634-252A-2 x US-08-477-407-2   ..

 Align seg 1/1  to: US-08-477-407-2  from: 1  to: 524

     502 CTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTAC...GA 548
         |||::::::   ||||||||||||:::|||||||||   |||:::   ::
       4 LeuSerHisGlnLeuLeuSerSerGlnTyrValGluArgHisPheSerAr 20

     549 AAATGGGAAACCACAGTACTCTAAGGGT...GGAGAGCACTGTTACTACC 595
         ::::|||      |||:::|||   |||   |||:::|||||||||||||:
      20 gGluGlyThrThrGlnHisSerThrGlyAlaGlyAspHisCysTyrTyrG 37

     596 ATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGC 645
         ::|||::::::||||||         |||   :::|||||||||||||||
      37 lnGlyLysLeuArgGlyAsnProHisSerPheAlaAlaLeuSerThrCys 53

     646 AATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGA 695
         :::|||||||||||:::|||:::|||   ::::::::::|||:::::||
      54 GlnGlyLeuHisGlyValPheSerAspGlyAsnLeuThrTyrIleValGl 70

     696 GCCACTAGAGCTGGTT......CATGATGAGAAAAGCACAGGTCGACCAC 739
         ||||   |||::::::                  :::         ||||
      70 uProGlnGluValAlaGlyProTrpGlyAlaProGlnGlyProLeuProH 87

     740 ATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAAT 789
         ||:::|||   :::|||
      87 isLeuIleTyrArgThr..............................  92

     790 CTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTG 839
                                 |||:::|||   :::
      93 ..........................ProLeuLeuProAspProLeuGl 100

     840 GTTGAAA.....................................  846
             :::
     100 yCysArgGluProGlyCysLeuPheAlaValProAlaGlnSerAlaProP 117

     847 ................AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT 879
                        ||||||||||||   |||:::      ||||||
     117 roAsnArgProArgLeuArgArgLysArgGlnValArg......ArgGly 131

     880 .........ATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAA 920
                 :::   :::|||   ||||||:::||||||:::::::::|||
     132 HisProThrValHisSerGluThrLysTyrValGluLeuIleValIleAs 148

     921 TGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAACA 970
         |||||||:::   :::::::::   |||:::|||   :::   |||:::|
     148 nAspHisGlnLeuPheGluGlnMetArgGlnSerValValLeuThrSerA 165

     971 ACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAG 1020
         ||||||||||||||||||||||||||:::|||   ||||||||||||||||
     165 snPheAlaLysSerValValAsnLeuAlaAspValIleTyrLysGluGln 181

    1021 CTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGA 1070
         ||||||||||||:::||||||||||||:::|||||||||::::::   ||
     182 LeuAsnThrArgIleValLeuValAlaMetGluThrTrpAlaAspGlyAs 198

    1071 TCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAA 1120
         |:::|||::::::      :::   ::::::   |||   ::::::
     198 pLysIleGlnValGlnAspAspLeuLeuGluThrLeuAlaArgLeuMetV 215

    1121 AATACCGGCAGCGC...ATTAAGCAGCATGCTGATGCTGTGCACCTCATC 1167
           ||||||:::::   :::   :::   ::::::|||:::||||||:::
     215 alTyrArgArgGluGlyLeuProGluProSerAsnAlaThrHisLeuPhe 231

    1168 TCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGG 1217
         |||      ||||||:::         |||:::   :::|||   |||||
     232 SerGlyArgThrPheGlnSerThrSerSerGlyAlaAlaTyrValGlyGl 248

    1218 TGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAA 1267
         |:::||||||   :::::::|||   |||||||||||||||||
     248 yIleCysSerLeuSerHisGlyGlyGlyValAsnGluTyrGlyAsnMetG 265

    1268 TGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATC 1317
           |||:::|||   :::|||:::|||:::|||:::|||||||||||||:::
     265 lyAlaMetAlaValThrLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMet 281

    1318 CAATGG...GAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATC 1364
         :::|||   :::      |||           |||   |||   :::
     282 MetTrpAsnLysHisArgSerSerAlaGlyAspCysLysCysProAspIl 298

    1365 CTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAAT 1414
          |||   |||||||||||||||:::|||||||            |||||||
     298 eTrpLeuGlyCysIleMetGluAspThrGlyPheTyrLeuProArgLysP 315

    1415 TTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGT 1464
         |||||:::|||||||||   |||||::::::|||||||||:::||||||
     315 heSerArgCysSerIleAspGluTyrAsnGlnPheLeuGlnGluGlyGly 331

    1465 GGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATG 1514
         |||:::|||||||||||||:::|||   ||||||:::::|||   |||||
     332 GlySerCysLeuPheAsnLysProLeuLysLeuLeuAspProProGluCy 348

    1515 TGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATG 1564
         ||||||||||:::|||||||||||||||||||||||||||||||||||||
     348 sGlyAsnGlyPheValGluAlaGlyGluGluCysAspCysGlySerValG 365

    1565 TGGAATGCTAT.........GGATTATGCTGTAAGAAATGTTCCCTCTCC 1605
           ||||||             |||   |||||||||||||||:::|||:::
     365 lnGluCysSerArgAlaGlyGlyAsnCysCysLysLysCysThrLeuThr 381

    1606 AACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCT 1655
         :::   |||   ||||||||||||   ||||||::::::      |||
     382 HisAspAlaMetCysSerAspGlyLeuCysCysArgArg......CysLy 396

    1656 TTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATA 1705
         ::::::|||||||||   :::|||||||:::|||||||||||||||||||
     396 sTyrGluProArgGlyValSerCysArgGluAlaValAsnGluCysAspI 413

    1706 TTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCAT 1755
         ||:::|||   |||||||||||||||:::|||||||||||||||||||||
     413 leAlaGluThrCysThrGlyAspSerSerGlnCysProProAsnLeuHis 429

    1756 AAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGG 1805
         |||   |||||||||   |||::::::::::|||||||||||||||:::||
     430 LysLeuAspGlyTyrTyrCysAspHisGluGlnGlyArgCysTyrGlyGl 446

    1806 CGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGG 1855
         |:::|||||||||||||||:::|||||||||   :::||||||      |
     446 yArgCysLysThrArgAspArgGlnCysGlnValLeuTrpGlyHisAlaA 463

    1856 CTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACT 1905
         |||||      |||:::|||||||||||||||||||||:::|||||||||
     463 laAla......AspArgPheCysTyrGluLysLeuAsnValGluGlyThr 477

    1906 GAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAA 1955
         |||:::|||:::||||||:::   |||:::   |||:::|||||||||||
     478 GluArgGlySerCysGlyArgLysGlySerGlyTrpValGlnCysSerLy 494

    1956 A 1956
         |
     494 s 494

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-484-355-2

seq_documentation_block:
; Sequence 2, Application US/08484355
; Patent No. 5705341
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  :  EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,355
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/243,542
;      FILING DATE:  13-MAY-1994
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
;      TELEFAX:  (616) 381-5465
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  524 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY:  human fetal brain cDNA library
US-08-484-355-2


alignment_scores:
            Quality: 1297.00              Length:    517
              Ratio:  3.515                Gaps:     12
 Percent Similarity:  71.373   Percent Identity: 49.516

alignment_block:
 US-09-634-252A-2 x US-08-484-355-2   ..

 Align seg 1/1  to: US-08-484-355-2  from: 1  to: 524

     502 CTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTAC...GA 548
         |||::::::   |||||||||||||:::|||||||||   |||:::   ::
       4 LeuSerHisGlnLeuLeuSerSerGlnTyrValGluArgHisPheSerAr 20

     549 AAATGGGAAACCACAGTACTCTAAGGGT...GGAGAGCACTGTTACTACC 595
         ::::|||      |||:::|||   |||   |||:::|||||||||||:
      20 gGluGlyThrThrGlnHisSerThrGlyAlaGlyAspHisCysTyrTyrG 37

     596 ATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGC 645
         ::|||::::::||||||         |||   :::|||||||||||||||
      37 lnGlyLysLeuArgGlyAsnProHisSerPheAlaAlaLeuSerThrCys 53

     646 AATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGA 695
         :::||||||||||||:::|||:::|||   ::::::::::|||:::::||
      54 GlnGlyLeuHisGlyValPheSerAspGlyAsnLeuThrTyrIleValGl 70

     696 GCCACTAGAGCTGGTT......CATGATGAGAAAAGCACAGGTCGACCAC 739
         ||||   |||:::::                    :::         ||||
      70 uProGlnGluValAlaGlyProTrpGlyAlaProGlnGlyProLeuProH 87

     740 ATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAAT 789
         ||:::|||   :::|||
      87 isLeuIleTyrArgThr..............................  92

     790 CTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTG 839
                                       |||:::|||   :::
      93 .........................ProLeuLeuProAspProLeuGl 100

     840 GTTGAAA.....................................  846
             :::
     100 yCysArgGluProGlyCysLeuPheAlaValProAlaGlnSerAlaProP 117

     847 ................AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT 879
                         ||||||||||||   |||:::      ||||||
     117 roAsnArgProArgLeuArgArgLysArgGlnValArg......ArgGly 131

     880 .........ATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAA 920
                 :::   :::|||   ||||||:::||||||:::::::::||
     132 HisProThrValHisSerGluThrLysTyrValGluLeuIleValIleAs 148

     921 TGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAACA 970
         |||||||:::   :::::::::   |||:::|||   :::   |||:::|
     148 nAspHisGlnLeuPheGluGlnMetArgGlnSerValValLeuThrSerA 165

     971 ACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAG 1020
         ||||||||||||||||||||||||:::|||   ||||||||||||||||
     165 snPheAlaLysSerValValAsnLeuAlaAspValIleTyrLysGluGln 181

    1021 CTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGA 1070
         ||||||||||||:::||||||||||||:::|||||||:::::::   ||
     182 LeuAsnThrArgIleValLeuValAlaMetGluThrTrpAlaAspGlyAs 198

    1071 TCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAA 1120
         |:::|||::::::      :::   ::::::   |||   ::::::
     198 pLysIleGlnValGlnAspAspLeuLeuGluThrLeuAlaArgLeuMetV 215

    1121 AATACCGGCAGCGC...ATTAAGCAGCATGCTGATGCTGTGCACCTCATC 1167
           ||||||::::::   :::   :::   ::::::|||:::||||||:::
     215 alTyrArgArgGluGlyLeuProGluProSerAsnAlaThrHisLeuPhe 231

    1168 TCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGG 1217
         |||      |||||:::         |||:::   :::|||   |||||
     232 SerGlyArgThrPheGlnSerThrSerSerGlyAlaAlaTyrValGlyGl 248

    1218 TGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAA 1267
         |:::||||||   ::::::|||   ||||||||||||||||||
     248 yIleCysSerLeuSerHisGlyGlyGlyValAsnGluTyrGlyAsnMetG 265

    1268 TGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATC 1317
           |||:::|||   :::|||:::|||:::|||:::|||||||||||:::
     265 lyAlaMetAlaValThrLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMet 281

    1318 CAATGG...GAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATC 1364
         :::|||   :::      |||            |||   |||   :::
     282 MetTrpAsnLysHisArgSerSerAlaGlyAspCysLysCysProAspIl 298

    1365 CTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAAT 1414
          |||   ||||||||||||||:::||||||            |||||||
     298 eTrpLeuGlyCysIleMetGluAspThrGlyPheTyrLeuProArgLysP 315

    1415 TTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGT 1464
         |||||:::|||||||||   ||||||:::::||||||||||:::|||||||
```

```
     315 heSerArgCysSerIleAspGluTyrAsnGlnPheLeuGlnGluGlyGly 331

    1465 GGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATG 1514
         |||:::|||||||||||||:::|||   ||||||:::::::|||   |||||
     332 GlySerCysLeuPheAsnLysProLeuLysLeuAspProProGluCy 348

    1515 TGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATG 1564
         ||||||||||:::||||||||||||||||||||||||||||
     348 sGlyAsnGlyPheValGluAlaGlyGluGluCysAspCysGlySerValG 365

    1565 TGGAATGCTAT.........GGATTATGCTGTAAGAAATGTTCCCTCTCC 1605
           ||||||            |||   |||||||||||||||:::|||:::
     365 lnGluCysSerArgAlaGlyGlyAsnCysCysLysLysCysThrLeuThr 381

    1606 AACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCT 1655
         :::   |||   ||||||||||||   ||||||:::::      |||
     382 HisAspAlaMetCysSerAspGlyLeuCysCysArgArg......CysLy 396

    1656 TTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATA 1705
          ::::::|||||||||   :::|||||||:::||||||||||||||||
     396 sTyrGluProArgGlyValSerCysArgGluAlaValAsnGluCysAspI 413

    1706 TTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCAT 1755
         ||:::|||   |||||||||||||||:::|||||||||||||||||||||
     413 leAlaGluThrCysThrGlyAspSerSerGlnCysProProAsnLeuHis 429

    1756 AAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGG 1805
         |||   |||||||||   |||:::::::::|||||||||||||||:::||
     430 LysLeuAspGlyTyrTyrCysAspHisGluGlnGlyArgCysTyrGlyGl 446

    1806 CGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGG 1855
         |:::|||||||||||||||:::|||||||||   :::||||||      |
     446 yArgCysLysThrArgAspArgGlnCysGlnValLeuTrpGlyHisAlaA 463

    1856 CTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACT 1905
         |||||      |||:::|||||||||||||||||||||:::||||||||
     463 laAla......AspArgPheCysTyrGluLysLeuAsnValGluGlyThr 477

    1906 GAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAA 1955
         |||:::|||:::||||||:::   |||:::   |||:::||||||||||
     478 GluArgGlySerCysGlyArgLysGlySerGlyTrpValGlnCysSerLy 494

    1956 A 1956
         |
     494 s 494

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-243-542-1

seq_documentation_block:
; Sequence 1, Application US/08243542
; Patent No. 5552526
;   GENERAL INFORMATION:
;     APPLICANT:  NAKAMURA, YUSUKE
;     APPLICANT:  :  EMI, MITSURU
;     TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;     TITLE OF INVENTION:  ENCODING THE SAME
;     NUMBER OF SEQUENCES:  20
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;       STREET:  2026 Rambling Road
;       CITY:  Kalamazoo
;       STATE:  Michigan
;       COUNTRY:  USA
;       ZIP:  49008-1699
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;       COMPUTER:  IBM PC/XT/AT Compatible
;       OPERATING SYSTEM:  MS-DOS 5.0
;       SOFTWARE:  WordPerfect 5.0
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/243,542
;       FILING DATE:
;       CLASSIFICATION:  530
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  JP 5-136602
;       FILING DATE:  14 MAY 1993
;       APPLICATION NUMBER:  JP 5-257455
;       FILING DATE:  22 SEPTEMBER 1993
;       APPLICATION NUMBER:  JP 6-49904
;       FILING DATE:  23 FEBRUARY 1994
;       APPLICATION NUMBER:  JP 6-73328
;       FILING DATE:  12 APRIL 1994
;       APPLICATION NUMBER:  JP 6-84470
;       FILING DATE:  22 APRIL 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Terryence F. Chapman
;       REGISTRATION NUMBER:  32 549
;       REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (616) 381-1156
;       TELEFAX:  (616) 381-5465
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  488 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
;     IMMEDIATE SOURCE:
;       LIBRARY:  human fetal brain cDNA library
US-08-243-542-1


alignment_scores:
            Quality: 1293.00              Length:    513
              Ratio:  3.533                 Gaps:     12
 Percent Similarity:  71.345   Percent Identity: 49.708

alignment_block:
 US-09-634-252A-2 x US-08-243-542-1   ..

 Align seg 1/1  to: US-08-243-542-1  from: 1  to: 488

     514 TTGTTGTCTTCTGATTATGTGGAGATTCACTAC...GAAAATGGGAAACC 560
         ||||||||||||:::|||||||||   |||:::   ::::::|||
       1 LeuLeuSerSerGlnTyrValGluArgHisPheSerArgGluGlyThrTh 17

     561 ACAGTACTCTAAGGGT...GGAGAGCACTGTTACTACCATGGAAGCATCA 607
          |||:::|||   |||   |||:::|||||||||||:::|||::::::|
      17 rGlnHisSerThrGlyAlaGlyAspHisCysTyrTyrGlnGlyLysLeuA 34

     608 GAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCAT 657
         |||||          |||   :::|||||||||||||||:::|||||||||
      34 rgGlyAsnProHisSerPheAlaAlaLeuSerThrCysGlnGlyLeuHis 50

     658 GGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCT 707
         |||:::|||:::|||   :::::::::|||::::::|||||   |||::
      51 GlyValPheSerAspGlyAsnLeuThrTyrIleValGluProGlnGluVa 67

     708 GGTT......CATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGA 751
         ::::                ::::         ||||||:::|||   :
      67 lAlaGlyProTrpGlyAlaProGlnGlyProLeuProHisLeuIleTyrA 84

     752 AAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAA 801
         ::|||
      84 rgThr..................................................... 85

     802 AGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAA..... 846
                     |||:::|||   :::            :::
      86 ..............ProLeuLeuProAspProLeuGlyCysArgGluPr 97

     846 ..................................................... 846

      97 oGlyCysLeuPheAlaValProAlaGlnSerAlaProProAsnArgProA 114
```

```
 847 .....AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT.........ATA 882
      ||||||||||||   |||:::      ||||||          :::
 114 rgLeuArgArgLysArgGlnValArg......ArgGlyHisProThrVal 128

 883 TTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAAC 932
     :::|||   ||||||:::|||||:::::::::|||||||||:::
 129 HisSerGluThrLysTyrValGluLeuIleValIleAsnAspHisGlnLe 145

 933 GTATAAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGT 982
      :::::::::   |||:::|||   :::   |||:::|||||||||||||
 145 uPheGluGlnMetArgGlnSerValValLeuThrSerAsnPheAlaLysS 162

 983 CCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGG 1032
     ||||||||||||||:::|||   |||||||||||||||||||||||| ||
 162 erValValAsnLeuAlaAspValIleTyrLysGluGlnLeuAsnThrArg 178

1033 GTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACAT 1082
     :::||||||||||||:::|||||||||:::::::   |||:::|||:::::
 179 IleValLeuValAlaMetGluThrTrpAlaAspGlyAspLysIleGlnVa 195

1083 CACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGC 1132
     :      :::   ::::::   |||   ::::::      ||||||:::::
 195 lGlnAspAspLeuLeuGluThrLeuAlaArgLeuMetValTyrArgArgG 212

1133 GC...ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACA 1179
     ::   :::   :::   ::::::|||:::|||||:::|||      |||
 212 luGlyLeuProGluProSerAsnAlaThrHisLeuPheSerGlyArgThr 228

1180 TTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCG 1229
     |||:::         |||:::   :::|||   |||||||:::||||||
 229 PheGlnSerThrSerSerGlyAlaAlaTyrValGlyGlyIleCysSerLe 245

1230 CACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCAC 1279
      ::::::|||   ||||||||||||||||||         |||:::|||
 245 uSerHisGlyGlyGlyValAsnGluTyrGlyAsnMetGlyAlaMetAlaV 262

1280 AAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGG...GAA 1326
     :::|||:::|||:::|||:::|||||||||||||:::::|||   :::
 262 alThrLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMetMetTrpAsnLys 278

1327 CCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTG 1376
          |||             |||   |||   :::   |||   |||||
 279 HisArgSerSerAlaGlyAspCysLysCysProAspIleTrpLeuGlyCy 295

1377 CATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCA 1426
     ||||||||||:::||||||           ||||||||||||:::||||
 295 sIleMetGluAspThrGlyPheTyrLeuProArgLysPheSerArgCysS 312

1427 GCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTT 1476
     |||||   ||||||:::::|||||||||:::||||||||||:::||||||
 312 erIleAspGluTyrAsnGlnPheLeuGlnGluGlyGlySerCysLeu 328

1477 TTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATA 1526
     ||||||:::|||   ||||||::::::|||   ||||||||||||||::
 329 PheAsnLysProLeuLysLeuLeuAspProProGluCysGlyAsnGlyPh 345

1527 CGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTAT. 1575
     :|||||||||||||||||||||||||||||         ||||||
 345 eValGluAlaGlyGluGluCysAspCysGlySerValGlnGluCysSerA 362

1576 ........GGATTATGCTGTAAGAAATGTTCCCTCTCCAACGGGGCTCAC 1617
             |||   ||||||||||||||||:::|||::::::   |||
 362 rgAlaGlyGlyAsnCysCysLysLysCysThrLeuThrHisAspAlaMet 378

1618 TGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACG 1667
     ||||||||||||   ||||||::::::       |||   ::::::|||||
 379 CysSerAspGlyLeuCysCysArgArg......CysLysTyrGluProAr 393

1668 AGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATT 1717
     ||||   :::||||||:::|||||||||||||||||||||:::|||   |
 393 gGlyValSerCysArgGluAlaValAsnGluCysAspIleAlaGluThrC 410

1718 GTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGA 1767
          ||||||||||||||:::|||||||||||||||||||||||   ||||||
      410 ysThrGlyAspSerSerGlnCysProProAsnLeuHisLysLeuAspGly 426

     1768 TATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGAC 1817
          |||   |||:::::::::|||||||||||||||||:::|||:::|||||||||
      427 TyrTyrCysAspHisGluGlnGlyArgCysTyrGlyGlyArgCysLysTh 443

     1818 CAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTG 1867
          |||||||:::|||||||||   :::||||||      ||||||      |
      443 rArgAspArgGlnCysGlnValLeuTrpGlyHisAlaAlaAla......A 458

     1868 ACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAAC 1917
          ||:::|||||||||||||||||||||:::|||||||||||:::|||:::
      458 spArgPheCysTyrGluLysLeuAsnValGluGlyThrGluArgGlySer 474

     1918 TGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAA 1956
          |||||||:::   |||:::   |||:::||||||||||||
      475 CysGlyArgLysGlySerGlyTrpValGlnCysSerLys 487
```

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-477-407-1

```
seq_documentation_block:
; Sequence 1, Application US/08477407
; Patent No. 5631351
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  :  EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE: . FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,407
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/243,542
;      FILING DATE:  13-MAY-1994
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
;      TELEFAX:  (616) 381-5465
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  488 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    ORIGINAL SOURCE:
```

```
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY:  human fetal brain cDNA library
US-08-477-407-1


alignment_scores:
            Quality: 1293.00              Length:    513
              Ratio:  3.533                Gaps:     12
 Percent Similarity:  71.345   Percent Identity: 49.708

alignment_block:
 US-09-634-252A-2 x US-08-477-407-1   ..

 Align seg 1/1  to: US-08-477-407-1  from: 1  to: 488

     514 TTGTTGTCTTCTGATTATGTGGAGATTCACTAC...GAAAATGGGAAACC 560
         ||||||||||||:::|||||||||   |||:::   :::::::|||
       1 LeuLeuSerSerGlnTyrValGluArgHisPheSerArgGluGlyThrTh 17

     561 ACAGTACTCTAAGGGT...GGAGAGCACTGTTACTACCATGGAAGCATCA 607
          |||:::|||   |||   |||:::||||||||||||:::|||::::::|
      17 rGlnHisSerThrGlyAlaGlyAspHisCysTyrTyrGlnGlyLysLeuA 34

     608 GAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCAT 657
         |||||         |||   :::|||||||||||||||:::|||||||||
      34 rgGlyAsnProHisSerPheAlaAlaLeuSerThrCysGlnGlyLeuHis 50

     658 GGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCT 707
         |||:::|||:::|||   ::::::::::|||:::::|||||   |||::
      51 GlyValPheSerAspGlyAsnLeuThrTyrIleValGluProGlnGluVa 67

     708 GGTT......CATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGA 751
         ::::                  :::         ||||||:::|||   :
      67 lAlaGlyProTrpGlyAlaProGlnGlyProLeuProHisLeuIleTyrA 84

     752 AAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAA 801
         ::|||
      84 rgThr..................................................... 85

     802 AGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAA..... 846
                       |||:::|||   :::             :::
      86 ..............ProLeuLeuProAspProLeuGlyCysArgGluPr 97

     846 ..................................................... 846

      97 oGlyCysLeuPheAlaValProAlaGlnSerAlaProProAsnArgProA 114

     847 .....AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT.........ATA 882
              ||||||||||||   |||:::      ||||||        :::
     114 rgLeuArgArgLysArgGlnValArg......ArgGlyHisProThrVal 128

     883 TTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAAC 932
            :::|||   ||||||:::||||||:::::::::|||||||||:::
     129 HisSerGluThrLysTyrValGluLeuIleValIleAsnAspHisGlnLe 145

     933 GTATAAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGT 982
          :::::::::   |||:::|||   :::   |||:::|||||||||||||
     145 uPheGluGlnMetArgGlnSerValValLeuThrSerAsnPheAlaLysS 162

     983 CCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGG 1032
         ||||||||||||||:::|||   ||||||||||||||||||||||||||
     162 erValValAsnLeuAlaAspValIleTyrLysGluGlnLeuAsnThrArg 178

    1033 GTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACAT 1082
         :::||||||||||||:::|||||||||::::::   |||:::|||:::::
     179 IleValLeuValAlaMetGluThrTrpAlaAspGlyAspLysIleGlnVa 195

    1083 CACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGC 1132
         :     :::   ::::::   |||   ::::::      |||||||::::
     195 lGlnAspAspLeuLeuGluThrLeuAlaArgLeuMetValTyrArgArgG 212

    1133 GC...ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACA 1179
         ::   :::   :::   ::::::|||:::|||||||:::|||      |||
     212 luGlyLeuProGluProSerAsnAlaThrHisLeuPheSerGlyArgThr 228

    1180 TTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCG 1229
         |||:::         |||:::   :::|||   ||||||:::||||||
     229 PheGlnSerThrSerSerGlyAlaAlaTyrValGlyGlyIleCysSerLe 245

    1230 CACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCAC 1279
          ::::::|||   ||||||||||||||||||         |||:::|||
     245 uSerHisGlyGlyGlyValAsnGluTyrGlyAsnMetGlyAlaMetAlaV 262

    1280 AAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGG...GAA 1326
           :::|||:::|||:::|||:::||||||||||||:::::::|||   :::
     262 alThrLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMetMetTrpAsnLys 278

    1327 CCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTG 1376
               |||            |||   |||   :::   |||   |||||
     279 HisArgSerSerAlaGlyAspCysLysCysProAspIleTrpLeuGlyCy 295

    1377 CATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCA 1426
         ||||||||||:::|||||            ||||||||||||:::||||
     295 sIleMetGluAspThrGlyPheTyrLeuProArgLysPheSerArgCysS 312

    1427 GCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTT 1476
         |||||   ||||||:::::|||||||||:::|||||||||:::||||||
     312 erIleAspGluTyrAsnGlnPheLeuGlnGluGlyGlySerCysLeu 328

    1477 TTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATA 1526
         ||||||:::|||   ||||||:::::::|||   ||||||||||||||::
     329 PheAsnLysProLeuLysLeuAspProProGluCysGlyAsnGlyPh 345

    1527 CGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTAT. 1575
         :|||||||||||||||||||||||||||||||         ||||||
     345 eValGluAlaGlyGluGluCysAspCysGlySerValGlnGluCysSerA 362

    1576 ........GGATTATGCTGTAAGAAATGTTCCCTCTCCAACGGGGCTCAC 1617
                 |||   ||||||||||||||:::|||::::::   |||
     362 rgAlaGlyGlyAsnCysCysLysLysCysThrLeuThrHisAspAlaMet 378

    1618 TGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACG 1667
         ||||||||||||   ||||||::::::      |||   ::::::|||||
     379 CysSerAspGlyLeuCysCysArgArg......CysLysTyrGluProAr 393

    1668 AGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATT 1717
         ||||   :::||||||:::||||||||||||||||||||||:::|||   |
     393 gGlyValSerCysArgGluAlaValAsnGluCysAspIleAlaGluThrC 410

    1718 GTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGA 1767
         ||||||||||||||:::||||||||||||||||||||||||   ||||||
     410 ysThrGlyAspSerSerGlnCysProProAsnLeuHisLysLeuAspGly 426

    1768 TATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGAC 1817
         |||   |||:::::::::|||||||||||||||||:::|||:::||||||||
     427 TyrTyrCysAspHisGluGlnGlyArgCysTyrGlyGlyArgCysLysTh 443

    1818 CAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTG 1867
         |||||||:::|||||||||   :::||||||      ||||||      |
     443 rArgAspArgGlnCysGlnValLeuTrpGlyHisAlaAlaAla......A 458

    1868 ACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAAC 1917
         ||:::|||||||||||||||||||:::||||||||||||:::|||:::
     458 spArgPheCysTyrGluLysLeuAsnValGluGlyThrGluArgGlySer 474

    1918 TGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAA 1956
         ||||||:::   |||:::   |||:::||||||||||||
     475 CysGlyArgLysGlySerGlyTrpValGlnCysSerLys 487

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-484-355-1

seq_documentation_block:
; Sequence 1, Application US/08484355
; Patent No. 5705341
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  : EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
       COMPUTER:  IBM PC/XT/AT Compatible
       OPERATING SYSTEM:  MS-DOS 5.0
       SOFTWARE:  WordPerfect 5.0
    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,355
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/243,542
;      FILING DATE:  13-MAY-1994
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
;      TELEFAX:  (616) 381-5465
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  488 amino acids
;      TYPE:  amino acid
       TOPOLOGY:  linear
    MOLECULE TYPE:  protein
    ORIGINAL SOURCE:
      ORGANISM:  Homo sapiens
    IMMEDIATE SOURCE:
;      LIBRARY:  human fetal brain cDNA library
US-08-484-355-1


alignment_scores:
            Quality: 1293.00              Length:    513
              Ratio:  3.533                Gaps:     12
 Percent Similarity:  71.345   Percent Identity: 49.708

alignment_block:
 US-09-634-252A-2 x US-08-484-355-1   ..

 Align seg 1/1  to: US-08-484-355-1  from: 1  to: 488

     514 TTGTTGTCTTCTGATTATGTGGAGATTCACTAC...GAAAATGGGAAACC 560
         ||||||||||||:::|||||||||   |||:::   ::::::|||
       1 LeuLeuSerSerGlnTyrValGluArgHisPheSerArgGluGlyThrTh 17

     561 ACAGTACTCTAAGGGT...GGAGAGCACTGTTACTACCATGGAAGCATCA 607
          |||:::|||   |||   |||:::|||||||||||||:::|||::::::|
      17 rGlnHisSerThrGlyAlaGlyAspHisCysTyrTyrGlnGlyLysLeuA 34

     608 GAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCAT 657
         |||||        |||   :::|||||||||||||:::|||||||||
      34 rgGlyAsnProHisSerPheAlaAlaLeuSerThrCysGlnGlyLeuHis 50

     658 GGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCT 707
         |||:::|||:::|||   ::::::::::|||:::::||||||   |||::
      51 GlyValPheSerAspGlyAsnLeuThrTyrIleValGluProGlnGluVa 67

     708 GGTT......CATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGA 751
         ::::                  :::         ||||||:::|||   :
      67 lAlaGlyProTrpGlyAlaProGlnGlyProLeuProHisLeuIleTyrA 84

     752 AAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAA 801
         ::|||
      84 rgThr..................................................... 85

     802 AGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAA..... 846
                     |||:::|||   :::           :::
      86 ..............ProLeuLeuProAspProLeuGlyCysArgGluPr 97

     846 ..................................................... 846

      97 oGlyCysLeuPheAlaValProAlaGlnSerAlaProProAsnArgProA 114

     847 .....AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT.........ATA 882
              ||||||||||||   |||:::      ||||||         :::
     114 rgLeuArgArgLysArgGlnValArg......ArgGlyHisProThrVal 128

     883 TTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAAC 932
         :::|||   ||||||:::||||||:::::::::|||||||||:::
     129 HisSerGluThrLysTyrValGluLeuIleValIleAsnAspHisGlnLe 145

     933 GTATAAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGT 982
          :::::::::   |||:::|||   :::   |||:::||||||||||||||
     145 uPheGluGlnMetArgGlnSerValValLeuThrSerAsnPheAlaLysS 162

     983 CCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGG 1032
         ||||||||||||||:::|||   ||||||||||||||||||||||||||||
     162 erValValAsnLeuAlaAspValIleTyrLysGluGlnLeuAsnThrArg 178

    1033 GTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACAT 1082
         :::|||||||||||||:::|||||||||::::::   |||:::|||:::::
     179 IleValLeuValAlaMetGluThrTrpAlaAspGlyAspLysIleGlnVa 195

    1083 CACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGC 1132
         :      :::   ::::::   |||   ::::::      ||||||:::
     195 lGlnAspAspLeuLeuGluThrLeuAlaArgLeuMetValTyrArgArgG 212

    1133 GC...ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACA 1179
         ::   :::   :::   ::::::|||:::||||||:::|||      |||
     212 luGlyLeuProGluProSerAsnAlaThrHisLeuPheSerGlyArgThr 228

    1180 TTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCG 1229
         |||:::         |||:::   :::|||   ||||||:::||||||
     229 PheGlnSerThrSerSerGlyAlaAlaTyrValGlyGlyIleCysSerLe 245

    1230 CACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCAC 1279
          ::::::|||   ||||||||||||||||||         |||:::|||
     245 uSerHisGlyGlyGlyValAsnGluTyrGlyAsnMetGlyAlaMetAlaV 262

    1280 AAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGG...GAA 1326
           :::|||:::|||:::|||:::|||||||||||||:::::|||   :::
     262 alThrLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMetMetTrpAsnLys 278

    1327 CCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTG 1376
         |||           |||   |||   :::   |||   |||||
     279 HisArgSerSerAlaGlyAspCysLysCysProAspIleTrpLeuGlyCy 295

    1377 CATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCA 1426
         ||||||||||:::||||||           ||||||||||||:::||||
     295 sIleMetGluAspThrGlyPheTyrLeuProArgLysPheSerArgCysS 312

    1427 GCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTT 1476
         |||||   ||||||:::::|||||||||:::|||||||||:::||||||
```

```
     312 erIleAspGluTyrAsnGlnPheLeuGlnGluGlyGlyGlySerCysLeu 328

    1477 TTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATA 1526
         ||||||:::|||   ||||||:::::::|||   ||||||||||||||::
     329 PheAsnLysProLeuLysLeuLeuAspProProGluCysGlyAsnGlyPh 345

    1527 CGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTAT. 1575
         :|||||||||||||||||||||||||||||||         ||||||
     345 eValGluAlaGlyGluGluCysAspCysGlySerValGlnGluCysSerA 362

    1576 ........GGATTATGCTGTAAGAAATGTTCCCTCTCCAACGGGGCTCAC 1617
                 |||   |||||||||||||||:::|||::::::   |||
     362 rgAlaGlyGlyAsnCysCysLysLysCysThrLeuThrHisAspAlaMet 378

    1618 TGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACG 1667
         ||||||||||||   ||||||::::::      |||   ::::::|||||
     379 CysSerAspGlyLeuCysCysArgArg......CysLysTyrGluProAr 393

    1668 AGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATT 1717
         ||||   :::||||||:::||||||||||||||||||||||:::|||   |
     393 gGlyValSerCysArgGluAlaValAsnGluCysAspIleAlaGluThrC 410

    1718 GTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGA 1767
         ||||||||||||||:::|||||||||||||||||||||||||   ||||||
     410 ysThrGlyAspSerSerGlnCysProProAsnLeuHisLysLeuAspGly 426

    1768 TATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGAC 1817
         |||   |||::::::::::|||||||||||||||:::|||:::|||||||
     427 TyrTyrCysAspHisGluGlnGlyArgCysTyrGlyGlyArgCysLysTh 443

    1818 CAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTG 1867
         |||||||:::|||||||||   :::||||||      ||||||      |
     443 rArgAspArgGlnCysGlnValLeuTrpGlyHisAlaAlaAla......A 458

    1868 ACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAAC 1917
         ||:::|||||||||||||||||||||:::|||||||||||:::|||:::
     458 spArgPheCysTyrGluLysLeuAsnValGluGlyThrGluArgGlySer 474

    1918 TGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAA 1956
         |||||||:::   |||:::   |||:::|||||||||||
     475 CysGlyArgLysGlySerGlyTrpValGlnCysSerLys 487

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-836-443-3

seq_documentation_block:
; Sequence 3, Application US/08836443
; Patent No. 5883241
;  GENERAL INFORMATION:
;    APPLICANT:  DOCHERTY, Andrew, J.P
;    APPLICANT:  SLOCOMBE, Patrick, M.
;    TITLE OF INVENTION:  DNA SEQUENCES CODING FOR A
;    TITLE OF INVENTION:  HUMAN METALLOPROTEINASE AND VARIANTS THEREOF
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,443
;      FILING DATE:  01-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  9709420
;      FILING DATE:  05-SEP-1997
;      APPLICATION NUMBER:  9612145.4
;      FILING DATE:  11-JUN-1996
;      APPLICATION NUMBER:  9526230.9
;      FILING DATE:  21-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Resnick, David S
;      REGISTRATION NUMBER:  34,235
;      REFERENCE/DOCKET NUMBER:  47424
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  751 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 5883241e
US-08-836-443-3


alignment_scores:
            Quality:  829.00            Length:    548
              Ratio:  2.375               Gaps:     15
 Percent Similarity:  63.686   Percent Identity: 33.394

alignment_block:
 US-09-634-252A-2 x US-08-836-443-3   ..

 Align seg 1/1  to: US-08-836-443-3  from: 1  to: 751

     835 CAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATT 884
         |||      |||:::|||   |||   :::      :::|||      ::
      17 GlnGlnThrLysLysArgProArgArgMet......LysArgGluAspLe 31

     885 TGAAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAACGT 934
         :::::::|||||||||:::||||||   :::|||   |||:::      :
      31 uAsnSerMetLysTyrValGluLeuTyrLeuValAlaAspTyrLeuGluP 48

     935 ATAAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCC 984
         :::::|||:::|||   ::::::      |||::::::          :::
      48 heGlnLysAsnArgArgAspGlnAspAlaThrLysHisLysLeuIleGlu 64

     985 GTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGT 1034
         ::::::|||   ||||||::::::|||:::   :::||||||   |||::
      65 IleAlaAsnTyrValAspLysPheTyrArg...SerLeuAsnIleArgIl 80

    1035 TGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCA 1084
         ::::||||||::::::|||:::||||||:::   ::::::   :::::::
      80 eAlaLeuValGlyLeuGluValTrpThrHisGlyAsnMetCysGluValS 97

    1085 CCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGC 1134
         ::   ||||||   :::   |||   :::|||   :::::::|||::::::
      97 erGluAsnProTyrSerThrLeuTrpSerPheLeuSerTrpArgArgLys 113

    1135 ATT......AAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGAC 1178
         :::      :::::::::   |||   ::::::||||||:::   :::::
      114 LeuLeuAlaGlnLysTyrHisAspAsnAlaGlnLeuIleThrGlyMetSe 130

    1179 ATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCT. 1227
         :||||||              :::|||:::   :::   ::::::||||||
      130 rPheHisGlyThrThrIleGlyLeuAlaProLeuMetAlaMetCysSerV 147

    1228 .....CGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCA 1272
               ::::::   ||||||:::::::::::::            ::::::
      147 alTyrGlnSerGlyGlyValAsnMetAspHisSerGluAsnAlaIleGly 163

    1273 GTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATG 1322
         ||||||   ::::::::::::::::::::::::|||:::|||:::
      164 ValAlaAlaThrMetAlaHisGluMetGlyHisAsnPheGlyMetThrHi 180

    1323 GGAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTG 1372
          :::   ::::::            |||:::::::::   :::   ||||
```

```
 180 sAspSerAlaAsp.........CysCysSerAlaSerAlaAlaAspGlyG 194

1373 GCTGCATCATG...GAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCA 1419
     ||||||||||          ||||||              :::   |||:::
 194 lyCysIleMetAlaAlaAlaThrGlyHisProPheProLysValPheAsn 210

1420 AAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGC 1469
        |||:::       |||          :::|||||||   |||||||||
 211 GlyCysAsnArgArgGluLeuAspArgTyrLeuGlnSerGlyGlyGlyMe 227

1470 CTGCCTTTTCAACAGGCCA......ACAAAGCTATTTGAGCCCACGGAAT 1513
      ||||||   |||   |||             |||:::          :::|
 227 tCysLeuSerAsnMetProAspThrArgMetLeuTyrGlyGlyArgArgC 244

1514 GTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCAT 1563
     ||||||||||||||:::|||   ||||||||||||||||||||    :::
 244 ysGlyAsnGlyTyrLeuGluAspGlyGluGluCysAspCysGlyGluGlu 260

1564 GTGGAATGCTATGGATTATGCTGT......AAGAAATGTTCCCTCTCCAA 1607
        ||||||   :::   ||||||      ::::::|||:::|||
 261 GluGluCysAsnAsnProCysCysAsnAlaSerAsnCysThrLeuArgPr 277

1608 CGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCCTT 1657
      ||||||:::|||:::   |||   ||||||:::      :::|||   :
 277 oGlyAlaGluCysAlaHisGlySerCysCysHis......GlnCysLysL 292

1658 TTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATT 1707
     ::         |||      ||||||:::   :::::::::||||||:::
 292 euLeuAlaProGlyThrLeuCysArgGluGlnAlaArgGlnCysAspLeu 308

1708 ACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCAAAA 1757
        |||:::||||||||||   |||   :::||||||   |||:::::::::
 309 ProGluPheCysThrGlyLysSerProHisCysProThrAsnPheTyrGl 325

1758 GCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCG 1807
     ::::||||||       |||:::   :::|||:::   ||||||||||||
 325 nMetAspGlyThrProCysGluGlyGlyGlnAlaTyrCysTyrAsnGlyM 342

1808 AGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCT 1857
        |||   |||   ::::::||||||||   :::||||||      |||
 342 etCysLeuThrTyrGlnGluGlnCysGlnGlnLeuTrpGlyProGlyAla 358

1858 GCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGA 1907
          :::      :::|||:::||||||:::|||:::   |||
 359 ArgProAlaProAspLeuCysPheGluLysValAsnValAlaGlyAspTh 375

1908 GAAGGGAAACTGCGGGAAGGAT...GGAGACCGGTGGATTCAGTGCAGCA 1954
        ||||||||||||||||||   :::   :::       :::|||:::
 375 rPheGlyAsnCysGlyLysAspMetAsnGlyGluHisArgLysCysAsnM 392

1955 AACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCT 2004
        :::|||:::   ||||||   :::   |||   :::
 392 etArgAspAlaLysCysGlyLysIleGlnCysGlnSerSerGluAlaArg 408

2005 CCACGTATTGGTCAACTTCAGGGTGAGATCATTCCA......ACTTCCTT 2048
     |||           |||:::::::::   :::|||      |||::::::
 409 Pro...........LeuGluSerAsnAlaValProIleAspThrThrIl 421

2049 CTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTA. 2097
     :      :::||||||   |||:::|||   |||:::||||||
 421 eIleMetAsnGlyArgGlnIleGlnCysArgGlyThrHisValTyrArgG 438

2098 ..............GATGATGATACGGATGTGGGCTATGTAGAAGATGGA 2133
                     |||       |||   |||   |||       |||
 438 lyProGluGluGluGlyAspMetLeuAspProGlyLeuValMetThrGly 454

2134 ACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAAT 2183
     |||   ||||||   :::   :::|||::::::   :::|||   :::
 455 ThrLysCysGlyTyrAsnHisIleCysPheGluGlyGlnCysArgAsnTh 471

2184 TCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCT 2233
      :::   ::::::   ::::::|||                |||||   |
 471 rSerPhePheGluThrGluGlyCys..............GlyLysLysC 483

     2234 GTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTC 2283
          ||:::|||||||||||||||:::|||:::   :::|||   |||
      483 ysAsnGlyHisGlyValCysAsnAsnAsnGlnAsnCysHisCysLeuPro 499

     2284 ACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCA 2333
               ||||||         |||:::                ::::::
      500 GlyTrpAlaProProPheCysAsnThrProGlyHisGlyGlySerIle.. 515

     2334 CCCCCCCAAGGATGAAGGACCCAAGGGTCCTAGTGCCACCAATCTCATAA 2383
                    |||:::||||||      |||:::::::::::   ::::
      516 .........AspSerGlyProMetProProGluSerValGlyProValV 529

     2384 TAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCTATTGTCCTT 2427
          ::::::::::       ||||||||||||   |||:::::::::
      529 alAlaGlyValLeuValAlaIleLeuValLeuAlaValLeuMet 543

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-765-243-8

seq_documentation_block:
; Sequence 8, Application US/08765243
; Patent No. 5935578
;  GENERAL INFORMATION:
;    APPLICANT:  ALVES, KENNETH
;    APPLICANT:  GUPTA, SUNIL K.
;    APPLICANT:  HOLLIS, GREGORY F.
;    TITLE OF INVENTION:  CONTRACEPTIVE VACCINE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MARY A. APPOLLINA
;      STREET:  P.O. BOX 2000, 126 E. LINCOLN AVENUE
;      CITY:  RAHWAY
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07065
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/765,243
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  APPOLLINA, MARY A
;      REGISTRATION NUMBER:  34,087
;      REFERENCE/DOCKET NUMBER:  19244Y
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)594-3462
;      TELEFAX:  (908)594-4720
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  734 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-765-243-8


alignment_scores:
            Quality:  739.00            Length:    695
              Ratio:  1.852               Gaps:     23
 Percent Similarity:  57.410   Percent Identity: 27.482

alignment_block:
 US-09-634-252A-2 x US-08-765-243-8   ..

 Align seg 1/1  to: US-08-765-243-8  from: 1  to: 734

     313 AATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACAT 362
         :::::::::       :::||||||:::|||:::::::::         ||
      20 AspSerLeuProValGlnIleThrValProGluLysIleArgSerIleIl 36
```

```
 363 CAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGAC 412
     |::::::      ||||||
  36 eLysGluGlyIleGluSer.............................. 42

 413 ACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAG 462
                                      |||||||||:::::::
  43 ................................GlnAlaSerTyrLys 47

 463 ATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGG 512
     |||         |||:::   :::   ::::::|||:::   :::|||
  48 IleValIleGluGlyLysProTyrThrValAsnLeuMetGlnLysAsnPh 64

 513 TTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAAATGGG.....A 556
      |||      :::      |||      :::      :::|||      |
  64 eLeuProHisAsnPheArgValTyrSerTyrSerGlyThrGlyIleMetL 81

 557 AACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATC 606
     |||||      :::         :::   |||:::|||:::|||   |||
  81 ysProLeuAspGlnAspPheGlnAsnPheCysHisTyrGlnGlyTyrIle 97

 607 AGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCA 656
     :::|||         |||   |||   :::|||||||||:::|||||::
  98 GluGlyTyrProLysSerValValMetValSerThrCysThrGlyLeuAr 114

 657 TGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGC 706
     :|||::::::::::   ::::::      |||   ||||||||||||
 114 gGlyValLeuGlnPheGluAsnValSerTyrGlyIleGluProLeu.... 129

 707 TGGTTCATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAACC 756
               |||:::|||:::|||      |||:::|||   ::::::
 130 ...........GluSerSerValGlyPheGluHisValIleTyrGlnVal 142

 757 TTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAGG 806
           ::::::::::|||         :::|||      |||:::
 143 .........LysHisLysLysAlaAspValSerLeuTyrAsnGluLysAs 156

 807 TGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAAGA 856
                         |||   :::   |||:::   :::   :
 156 p..................IleGluSerArgAspLeuSerPheLysLeuG 167

 857 GAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATTTGGAA 906
     ::::::::::|||::::::         :::      ||||||:::|||
 167 lnSerAlaGluProGlnGln.........AspPheAlaLysTyrIleGlu 180

 907 CTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCA 956
     :::   ::::::   :::   :::   |||:::         |||:::
 181 MetHisValIleValGluLysGlnLeuTyrAsnHisMetGlySerAspTh 197

 957 TGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTA 1006
      :::   :::   :::   :::   :::::::::::|||:::::::::|
 197 rThrValValAlaGlnLysValPheGlnLeuIleGlyLeuThrAsnAlaI 214

 007 TTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACC 1056
     ||:::      ::::::|||      ::::::|||   ::::::|||
 214 lePhe...ValSerPheAsnIleThrIleIleLeuSerSerLeuGluLeu 229

1057 TGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAGATGCT 1106
     |||   ::::::::::::|||      |||   :::      ::::::||
 230 TrpIleAspGluAsnLysIleAlaThrThrGlyGluAlaAsnGluLeuLe 246

1107 CCATGAGTTCTCAAAATACCGGCAGCGC.........ATTAAGCAGCATG 1147
     ||||   |||   :::::::::                 ::::::   |||
 246 uHisThrPheLeuArgTrpLysThrSerTyrLeuValLeuArgProHisA 263

1148 CTGATGCTGTGCACCTCATCTCGCGG.....................GTG 1176
          |||      |||:::   |||                     :::
 263 spValAlaPheLeuLeuValTyrArgGluLysSerAsnTyrValGlyAla 279

1177 ACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTC 1226
     ||||||:::   |||      :::   :::|||   |||||||||
 280 ThrPheGlnGlyLysMetCysAspAlaAsnTyrAlaGlyGlyValValLe 296
```

```
1227 TCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGG 1276
      :::   |||   ::::::::::::                    ::::::|
 296 uHisProArgThrIleSerLeuGlu.................SerLeuA 307

1277 CACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAA 1326
     ||   :::|||:::|||   |||:::   ::::::||||||   ::::::
 307 laValIleLeuAlaGlnLeuLeuSerLeuSerMetGlyIleThrTyrAsp 323

1327 CCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTG 1376
           :::          ||||||:::|||      |||   :::   ||
 324 AspIleAsn.........LysCysGlnCys......SerGlyAlaValCy 335

1377 CATCATGGAGGAAACAGGGGTGTCCCATTCT......CGAAAATTTTCAA 1420
     |||||||:::      ::::::      |||      :::   ||||||:
 335 sIleMetAsnProGluAlaIleHisPheSerGlyValLysIlePheSerA 352

1421 AGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCC 1470
     ::||||||:::   ::::::      |||:::::::::      :::
 352 snCysSerPheGluAspPheAlaHisPheIleSerLysGlnLysSerGln 368

1471 TGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCC.............. 1506
     ||||||   |||:::|||         :::::::|||
 369 CysLeuHisAsnGlnProArg......LeuAspProPhePheLysGlnGl 383

1507 .ACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTG 1555
      :::   |||||||||:::   :::|||||||||||||||||||||||||
 383 nAlaValCysGlyAsnAlaLysLeuGluAlaGlyGluGluCysAspCysG 400

1556 GTTTTCATGTGGAATGCTATGGATTA.........TGCTGT......AAG 1590
     ||   :::   :::|||      :::      ||||||
 400 lyThrGluGlnAspCysAlaLeuIleGlyGluThrCysCysAspIleAla 416

1591 AAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAA 1640
        |||   ::::::   |||:::::::|||::::::|||||||||||::
 417 ThrCysArgPheLysAlaGlySerAsnCysAlaGluGlyProCysCysGl 433

1641 CAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTG 1690
     :|||      |||||||||:::   :::         ||||||   :::
 433 uAsn......CysLeuPheMetSerLysGluArgMetCysArgProSerP 448

1691 TGAACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGC 1740
       :::|||||||||:::   |||||||||:::|||:::|||:::::::|||
 448 heGluGluCysAspLeuProGluTyrCysAsnGlySerSerAlaSerCys 464

1741 CCACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGG 1790
     |||   |||   :::   |||   |||:::   |||:::   ||||||
 465 ProGluAsnHisTyrValGlnThrGlyHisProCysGlyLeuAsnGlnTr 481

1791 CCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACA 1840
        |||   :::|||   |||   :::   |||:::||||||
 481 pIleCysIleAspGlyValCysMetSerGlyAspLysGlnCysThrAspT 498

1841 TCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTG 1890
       :::|||   ::::::      :::   :::   ||||||:::   |||
 498 hrPheGlyLysGluValGluPheGlyProSerGluCysTyrSerHisLeu 514

1891 AATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTG 1940
     |||:::::::          :::||||||||||||   :::   :::   ::
 515 AsnSerLysThrAspValSerGlyAsnCysGlyIleSerAspSerGlyTy 531

1941 GATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCA 1990
     :   ||||||:::      ::::::   ||||||   |||:::|||
 531 rThrGlnCysGluAlaAspAsnLeuGlnCysGlyLysLeuIleCysLysT 548

1991 AT............CTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGT 2028
                    |||   :::   ||||||         :::   :::
 548 yrValGlyLysPheLeuLeuGlnIleProArgAlaThrIleIleTyrAla 564

2029 GAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAG 2078
     :::|||                  |||         :::::::::   ::
 565 AsnIleSerGly............HisLeuCysIleAlaValGluPheAl 577

2079 TGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTATGTAGAAG 2128
```

```
      ::::   |||:::            |||::::::      :::::::::|
   577 aSerAspHisAla............AspSerGlnLysMetTrpIleLysA 590

  2129 ATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTA 2178
       ||||||||   ||||||   :::   :::|||   ::::::::::|||:::
   590 spGlyThrSerCysGlySerAsnLysValCysArgAsnGlnArgCysVal 606

  2179 CAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAA 2228
       :::   :::   |||:::      :::|||      |||:::
   607 SerSerSerTyrLeuGlyTyr...AspCysThrThrAspLys........ 619

  2229 AGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTG 2278
           |||:::   :::||||||||||:::|||:::      |||   |||:
   620 ....CysAsnAspArgGlyValCysAsnAsnLysLysHisCysHisCysS 635

   279 ATTTCACCTGGGCAGGGACAGATTGCAGTATCCGG 2313
       ::   ::::::          |||||||||:::::
   635 erAlaSerTyrLeuProProAspCysSerValGln 646

seq_name: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-07295-8

seq_documentation_block:
; Sequence 8, Application PC/TUS9507295
;  GENERAL INFORMATION:
;    APPLICANT:  ALVES, KENNETH
;    APPLICANT:  GUPTA, SUNIL K.
;    APPLICANT:  HOLLIS, GREGORY F.
;    TITLE OF INVENTION:  CONTRACEPTIVE VACCINE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MARY A. APPOLLINA
;      STREET:  P.O. BOX 2000, 126 E. LINCOLN AVENUE
;      CITY:  RAHWAY
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07065
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/07295
;      FILING DATE:
;      CLASSIFICATION:
     ATTORNEY/AGENT INFORMATION:
       NAME:  APPOLLINA, MARY A
       REGISTRATION NUMBER:  34,087
       REFERENCE/DOCKET NUMBER:  19244Y
     TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)594-3462
;      TELEFAX:  (908)594-4720
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  734 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US95-07295-8


alignment_scores:
            Quality:  739.00              Length:    695
              Ratio:  1.852                Gaps:     23
 Percent Similarity:  57.410   Percent Identity: 27.482

alignment_block:
 US-09-634-252A-2 x PCT-US95-07295-8   ..

 Align seg 1/1  to: PCT-US95-07295-8  from: 1  to: 734

   313 AATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACAT 362
       :::::::::      :::||||||:::|||:::::::::        ||

    20 AspSerLeuProValGlnIleThrValProGluLysIleArgSerIleIl 36

   363 CAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGAC 412
       |::::::      ||||||
    36 eLysGluGlyIleGluSer............................... 42

   413 ACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAG 462
                                        |||||||||::::::
    43 .................................GlnAlaSerTyrLys 47

   463 ATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGG 512
       |||          |||:::   :::   ::::::|||:::   :::|||
    48 IleValIleGluGlyLysProTyrThrValAsnLeuMetGlnLysAsnPh 64

   513 TTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAAATGGG......A 556
        |||      :::      |||      :::      :::|||      |
    64 eLeuProHisAsnPheArgValTyrSerTyrSerGlyThrGlyIleMetL 81

   557 AACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATC 606
       |||||      :::          :::   |||:::|||:::|||   |||
    81 ysProLeuAspGlnAspPheGlnAsnPheCysHisTyrGlnGlyTyrIle 97

   607 AGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCA 656
       :::|||         |||   |||   :::|||||||||:::||||||::
    98 GluGlyTyrProLysSerValValMetValSerThrCysThrGlyLeuAr 114

   657 TGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGC 706
       :|||::::::::::  ::::::      |||   ||||||||||||
   114 gGlyValLeuGlnPheGluAsnValSerTyrGlyIleGluProLeu.... 129

   707 TGGTTCATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAACC 756
                |||:::|||:::|||      |||:::|||   ::::::
   130 ...........GluSerServalGlyPheGluHisValIleTyrGlnVal 142

   757 TTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAGG 806
             :::::::::|||         :::|||      |||:::
   143 .........LysHisLysLysAlaAspValSerLeuTyrAsnGluLysAs 156

   807 TGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAAGA 856
                              |||   :::   |||:::   :::   :
   156 p.................IleGluSerArgAspLeuSerPheLysLeuG 167

   857 GAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATTTGGAA 906
       ::::::::::::|||::::::         :::      ||||||:::|||
   167 lnSerAlaGluProGlnGln.........AspPheAlaLysTyrIleGlu 180

   907 CTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCA 956
       :::   ::::::   :::   :::   |||:::         |||:::
   181 MetHisValIleValGluLysGlnLeuTyrAsnHisMetGlySerAspTh 197

   957 TGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTA 1006
        :::   :::   :::   :::   ::::::::::::|||:::::::::|
   197 rThrValValAlaGlnLysValPheGlnLeuIleGlyLeuThrAsnAlaI 214

  1007 TTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACC 1056
       ||:::      ::::::|||      :::::|||   ::::::|||
   214 lePhe...ValSerPheAsnIleThrIleIleLeuSerSerLeuGluLeu 229

  1057 TGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAGATGCT 1106
       |||   ::::::::::::|||      |||   :::      ::::::||
   230 TrpIleAspGluAsnLysIleAlaThrThrGlyGluAlaAsnGluLeuLe 246

  1107 CCATGAGTTCTCAAAATACCGGCAGCGC.........ATTAAGCAGCATG 1147
       ||||   |||   :::::::::                ::::::   |||
   246 uHisThrPheLeuArgTrpLysThrSerTyrLeuValLeuArgProHisA 263

  1148 CTGATGCTGTGCACCTCATCTCGCGG.....................GTG 1176
            |||      |||:::   |||                     :::
   263 spValAlaPheLeuLeuValTyrArgGluLysSerAsnTyrValGlyAla 279

  1177 ACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTC 1226
       ||||||:::   |||      :::   :::|||   |||||||||
   280 ThrPheGlnGlyLysMetCysAspAlaAsnTyrAlaGlyGlyValValLe 296
```

```
1227 TCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGG 1276
      :::   |||   ::::::::::::                    ::::::|
 296 uHisProArgThrIleSerLeuGlu.................SerLeuA 307

1277 CACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAA 1326
     ||   :::|||:::|||   |||:::   ::::::||||||   ::::::
 307 laValIleLeuAlaGlnLeuLeuSerLeuSerMetGlyIleThrTyrAsp 323

1327 CCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTG 1376
           :::         ||||||:::|||      |||   :::   ||
 324 AspIleAsn.........LysCysGlnCys......SerGlyAlaValCy 335

1377 CATCATGGAGGAAACAGGGGTGTCCCATTCT......CGAAAATTTTCAA 1420
     |||||||:::     ::::::     |||      :::   ||||||:
 335 sIleMetAsnProGluAlaIleHisPheSerGlyValLysIlePheSerA 352

1421 AGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCC 1470
     ::||||||:::   ::::::      |||::::::::::      :::
 352 snCysSerPheGluAspPheAlaHisPheIleSerLysGlnLysSerGln 368

1471 TGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCC.............. 1506
     ||||||   |||:::|||         ::::::|||
 369 CysLeuHisAsnGlnProArg......LeuAspProPhePheLysGlnGl 383

1507 .ACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTG 1555
      :::   |||||||||:::   :::|||||||||||||||||||||||||
 383 nAlaValCysGlyAsnAlaLysLeuGluAlaGlyGluGluCysAspCysG 400

1556 GTTTTCATGTGGAATGCTATGGATTA.........TGCTGT......AAG 1590
     ||   :::   :::|||      :::        ||||||
 400 lyThrGluGlnAspCysAlaLeuIleGlyGluThrCysCysAspIleAla 416

1591 AAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAA 1640
        |||   ::::::   |||:::::::|||::::::|||||||||||::
 417 ThrCysArgPheLysAlaGlySerAsnCysAlaGluGlyProCysCysGl 433

1641 CAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTG 1690
     :|||      |||||||||:::   :::         ||||||   :::
 433 uAsn......CysLeuPheMetSerLysGluArgMetCysArgProSerP 448

1691 TGAACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGC 1740
       :::|||||||||:::   |||||||||:::|||:::|||:::::::|||
 448 heGluGluCysAspLeuProGluTyrCysAsnGlySerSerAlaSerCys 464

1741 CCACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGG 1790
     |||   |||   :::   |||   |||:::   |||:::   ||||||
 465 ProGluAsnHisTyrValGlnThrGlyHisProCysGlyLeuAsnGlnTr 481

1791 CCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACA 1840
        |||   :::|||   |||   :::   |||:::||||||
 481 pIleCysIleAspGlyValCysMetSerGlyAspLysGlnCysThrAspT 498

841 TCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTG 1890
      :::|||   ::::::      :::   :::   ||||||:::   |||
 498 hrPheGlyLysGluValGluPheGlyProSerGluCysTyrSerHisLeu 514

1891 AATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTG 1940
     |||::::::          :::||||||||||||   :::   :::   ::
 515 AsnSerLysThrAspValSerGlyAsnCysGlyIleSerAspSerGlyTy 531

1941 GATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCA 1990
     :   ||||||:::      ::::::   ||||||   |||:::|||
 531 rThrGlnCysGluAlaAspAsnLeuGlnCysGlyLysLeuIleCysLysT 548

1991 AT...........CTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGT 2028
                |||   :::   ||||||         :::   :::
 548 yrValGlyLysPheLeuLeuGlnIleProArgAlaThrIleIleTyrAla 564

2029 GAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAG 2078
     :::|||                |||         :::::::::   ::
 565 AsnIleSerGly...........HisLeuCysIleAlaValGluPheAl 577

2079 TGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTATGTAGAAG 2128
          ::::   |||:::            |||::::::      :::::::::|
  577 aSerAspHisAla...........AspSerGlnLysMetTrpIleLysA 590

2129 ATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTA 2178
     ||||||||   ||||||   :::   :::|||   ::::::::::|||:::
  590 spGlyThrSerCysGlySerAsnLysValCysArgAsnGlnArgCysVal 606

2179 CAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAA 2228
          :::   :::   |||:::      :::|||      |||:::
  607 SerSerSerTyrLeuGlyTyr...AspCysThrThrAspLys........ 619

2229 AGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTG 2278
          |||:::   :::||||||||||:::|||:::      |||   |||:
  620 ....CysAsnAspArgGlyValCysAsnAsnLysLysHisCysHisCysS 635

2279 ATTTCACCTGGGCAGGGACAGATTGCAGTATCCGG 2313
     ::   ::::::          |||||||||::::::
  635 erAlaSerTyrLeuProProAspCysSerValGln 646
```

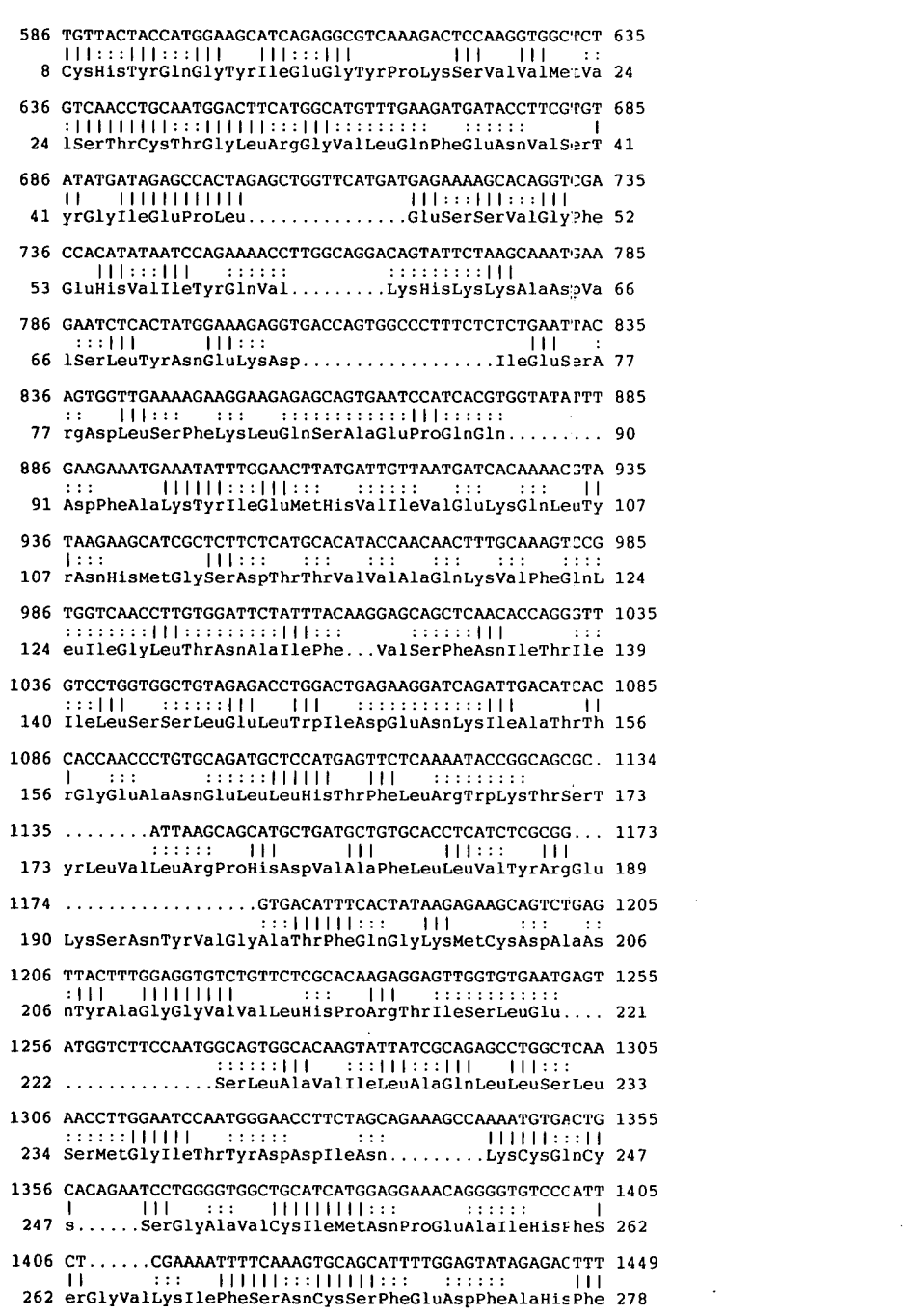
```
seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-264-101-2

seq_documentation_block:
; Sequence 2, Application US/08264101
; Patent No. 5693496
;  GENERAL INFORMATION:
;    APPLICANT:  ALVES, KENNETH
;    APPLICANT:  GUPTA, SUNIL K.
;    APPLICANT:  HOLLIS, GREGORY F.
;    TITLE OF INVENTION:  CONTRACEPTIVE VACCINE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MARY A. APPOLLINA
;      STREET:  P.O. BOX 2000, 126 E. LINCOLN AVENUE
;      CITY:  RAHWAY
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07065
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/264,101
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  APPOLLINA, MARY A.
;      REGISTRATION NUMBER:  34,087
;      REFERENCE/DOCKET NUMBER:  19244
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)594-3462
;      TELEFAX:  (908)594-4720
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  651 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-264-101-2


alignment_scores:
           Quality:  711.00              Length:    602
             Ratio:  1.986                 Gaps:     21
 Percent Similarity:  59.468   Percent Identity: 28.738

alignment_block:
 US-09-634-252A-2 x US-08-264-101-2   ..

 Align seg 1/1  to: US-08-264-101-2  from: 1  to: 651
```


586 TGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCT 635
8 CysHisTyrGlnGlyTyrIleGluGlyTyrProLysSerValValMetVa 24
636 GTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGT 685
24 lSerThrCysThrGlyLeuArgGlyValLeuGlnPheGluAsnValSerT 41
686 ATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACAGGTCGA 735
41 yrGlyIleGluProLeu.....GluSerSerValGlyPhe 52
736 CCACATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAA 785
53 GluHisValIleTyrGlnVal.....LysHisLysLysAlaAspVa 66
786 GAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTAC 835
66 lSerLeuTyrAsnGluLysAsp.....IleGluSerA 77
836 AGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATTT 885
77 rgAspLeuSerPheLysLeuGlnSerAlaGluProGlnGln..... 90
886 GAAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTA 935
91 AspPheAlaLysTyrIleGluMetHisValIleValGluLysGlnLeuTy 107
936 TAAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCG 985
107 rAsnHisMetGlySerAspThrThrValValAlaGlnLysValPheGlnL 124
986 TGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTT 1035
124 euIleGlyLeuThrAsnAlaIlePhe...ValSerPheAsnIleThrIle 139
1036 GTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCAC 1085
140 IleLeuSerSerLeuGluLeuTrpIleAspGluAsnLysIleAlaThrTh 156
1086 CACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGC. 1134
156 rGlyGluAlaAsnGluLeuLeuHisThrPheLeuArgTrpLysThrSerT 173
1135 ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGG... 1173
173 yrLeuValLeuArgProHisAspValAlaPheLeuLeuValTyrArgGlu 189
1174 GTGACATTTCACTATAAGAGAAGCAGTCTGAG 1205
190 LysSerAsnTyrValGlyAlaThrPheGlnGlyLysMetCysAspAlaAs 206
1206 TTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGT 1255
206 nTyrAlaGlyGlyValValLeuHisProArgThrIleSerLeuGlu.... 221
1256 ATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAA 1305
222 SerLeuAlaValIleLeuAlaGlnLeuLeuSerLeu 233
1306 AACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATGTGACTG 1355
234 SerMetGlyIleThrTyrAspAspIleAsn.....LysCysGlnCy 247
1356 CACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCATT 1405
247 s.....SerGlyAlaValCysIleMetAsnProGluAlaIleHisPheS 262
1406 CT.....CGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTT 1449
262 erGlyValLysIlePheSerAsnCysSerPheGluAspPheAlaHisPhe 278
1450 TTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATT 1499
279 IleSerLysGlnLysSerGlnCysLeuHisAsnGlnProArg.....Le 293
1500 TGAGCCC.....ACGGAATGTGGAAATGGATACGTGGAAG 1534
293 uAspProPhePheLysGlnGlnAlaValCysGlyAsnAlaLysLeuGluA 310
1535 CTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTATGGATTA... 1581
310 laGlyGluGluCysAspCysGlyThrGluGlnAspCysAlaLeuIleGly 326
1582 TGCTGT.....AAGAAATGTTCCCTCTCCAACGGGGCTCACTG 1619
327 GluThrCysCysAspIleAlaThrCysArgPheLysAlaGlySerAsnCy 343
1620 CAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAG 1669
343 sAlaGluGlyProCysCysGluAsn.....CysLeuPheMetSerLysG 358
1670 GGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGT 1719
358 luArgMetCysArgProSerPheGluGluCysAspLeuProGluTyrCys 374
1720 ACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATA 1769
375 AsnGlySerSerAlaSerCysProGluAsnHisTyrValGlnThrGlyHi 391
1770 TGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCA 1819
391 sProCysGlyLeuAsnGlnTrpIleCysIleAspGlyValCysMetSerG 408
1820 GAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGAC 1869
408 lyAspLysGlnCysThrAspThrPheGlyLysGluValGluPheGlyPro 424
1870 AAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTG 1919
425 SerGluCysTyrSerHisLeuAsnSerLysThrAspValSerGlyAsnCy 441
1920 CGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCT 1969
441 sGlyIleSerAspSerGlyTyrThrGlnCysGluAlaAspAsnLeuGlnC 458
1970 GTGGATTCTTACTCTGTACCAAT.....CTTACTCGAGCTCCA 2007
458 ysGlyLysLeuIleCysLysTyrValGlyLysPheLeuLeuGlnIlePro 474
2008 CGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCA 2057
475 ArgAlaThrIleIleTyrAlaAsnIleSerGly.....HisLe 487
2058 AGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATA 2107
487 uCysIleAlaValGluPheAlaSerAspHisAla.....AspS 500
2108 CGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATG 2157
500 erGlnLysMetTrpIleLysAspGlyThrSerCysGlySerAsnLysVal 516
2158 TGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTG 2207
517 CysArgAsnGlnArgCysValSerSerSerTyrLeuGlyTyr...AspCy 532
2208 TCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTA 2257
532 sThrThrAspLys.....CysAsnAspArgGlyValCysAsnA 545
2258 ATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGT 2307
545 snLysLysHisCysHisCysSerAlaSerTyrLeuProProAspCysSer 561
2308 ATCCGG 2313

```
      ::::::
    562 ValGln 563

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-765-243-2

seq_documentation_block:
; Sequence 2, Application US/08765243
; Patent No. 5935578
;  GENERAL INFORMATION:
;    APPLICANT:  ALVES, KENNETH
;    APPLICANT:  GUPTA, SUNIL K.
;    APPLICANT:  HOLLIS, GREGORY F.
;    TITLE OF INVENTION:  CONTRACEPTIVE VACCINE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MARY A. APPOLLINA
;      STREET:  P.O. BOX 2000, 126 E. LINCOLN AVENUE
;      CITY:  RAHWAY
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07065
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/765,243
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  APPOLLINA, MARY A
;      REGISTRATION NUMBER:  34,087
;      REFERENCE/DOCKET NUMBER:  19244Y
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)594-3462
;      TELEFAX:  (908)594-4720
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  651 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-765-243-2


alignment_scores:
            Quality:  711.00              Length:    602
              Ratio:  1.986                Gaps:     21
 Percent Similarity:  59.468   Percent Identity: 28.738

 [redacted]ment_block:
 [redacted]9-634-252A-2 x US-08-765-243-2   ..

 Align seg 1/1  to: US-08-765-243-2  from: 1  to: 651

     586 TGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCT 635
         |||:::|||:::|||   |||:::|||         |||   |||   ::
       8 CysHisTyrGlnGlyTyrIleGluGlyTyrProLysSerValValMetVa 24

     636 GTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGT 685
         :|||||||||:::|||||:::|||:::::::::   ::::::      |
      24 lSerThrCysThrGlyLeuArgGlyValLeuGlnPheGluAsnValSerT 41

     686 ATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACAGGTCGA 735
         ||   ||||||||||||              |||:::|||:::|||
      41 yrGlyIleGluProLeu.............GluSerServalGlyPhe 52

     736 CCACATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAA 785
         |||:::|||   :::::         ::::::::::|||
      53 GluHisValIleTyrGlnVal.........LysHisLysLysAlaAspVa 66

     786 GAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTAC 835
         :::|||      |||:::                          |||   :
      66 lSerLeuTyrAsnGluLysAsp.................IleGluSerA 77

     836 AGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATTT 885
         ::   |||:::   :::   ::::::::::::|||::::::
      77 rgAspLeuSerPheLysLeuGlnSerAlaGluProGlnGln......... 90

     886 GAAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTA 935
         :::      ||||||:::|||:::   ::::::   :::   :::   ||
      91 AspPheAlaLysTyrIleGluMetHisValIleValGluLysGlnLeuTy 107

     936 TAAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCG 985
         |.::         |||:::   :::   :::   :::   :::   ::::
     107 rAsnHisMetGlySerAspThrThrValValAlaGlnLysValPheGlnL 124

     986 TGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTT 1035
         ::::::::|||:::::::::|||:::      ::::::|||      :::
     124 euIleGlyLeuThrAsnAlaIlePhe...ValSerPheAsnIleThrIle 139

    1036 GTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCAC 1085
         :::|||   ::::::|||   |||   ::::::::::::|||      ||
     140 IleLeuSerSerLeuGluLeuTrpIleAspGluAsnLysIleAlaThrTh 156

    1086 CACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGC. 1134
         |   ::      ::::::||||||   |||   :::::::::
     156 rGlyGluAlaAsnGluLeuLeuHisThrPheLeuArgTrpLysThrSerT 173

    1135 ........ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGG... 1173
                 ::::::   |||      |||      |||:::   |||
     173 yrLeuValLeuArgProHisAspValAlaPheLeuLeuValTyrArgGlu 189

    1174 ..................GTGACATTTCACTATAAGAGAAGCAGTCTGAG 1205
                           :::|||||||:::   |||      :::   ::
     190 LysSerAsnTyrValGlyAlaThrPheGlnGlyLysMetCysAspAlaAs 206

    1206 TTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGT 1255
         :|||   |||||||||      :::   |||   ::::::::::::
     206 nTyrAlaGlyGlyValValLeuHisProArgThrIleSerLeuGlu.... 221

    1256 ATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAA 1305
                      :::::::|||   :::|||:::|||   |||:::
     222 ..............SerLeuAlaValIleLeuAlaGlnLeuLeuSerLeu 233

    1306 AACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATGTGACTG 1355
         ::::::||||||   ::::::      :::          ||||||:::||
     234 SerMetGlyIleThrTyrAspAspIleAsn.........LysCysGlnCy 247

    1356 CACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCATT 1405
         |      |||   :::   ||||||||:::      ::::::      |
     247 s......SerGlyAlaValCysIleMetAsnProGluAlaIleHisPheS 262

    1406 CT......CGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTT 1449
         ||      :::   ||||||:::||||||:::   ::::::      |||
     262 erGlyValLysIlePheSerAsnCysSerPheGluAspPheAlaHisPhe 278

    1450 TTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATT 1499
         :::::::::       :::   ||||||   |||:::|||         ::
     279 IleSerLysGlnLysSerGlnCysLeuHisAsnGlnProArg......Le 293

    1500 TGAGCCC..............ACGGAATGTGGAAATGGATACGTGGAAG 1534
         ::::|||               :::   ||||||||:::   :::||||
     293 uAspProPhePheLysGlnGlnAlaValCysGlyAsnAlaLysLeuGluA 310

    1535 CTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTATGGATTA... 1581
         ||||||||||||||||||||||||   :::   :::|||      :::
     310 laGlyGluGluCysAspCysGlyThrGluGlnAspCysAlaLeuIleGly 326

    1582 ......TGCTGT......AAGAAATGTTCCCTCTCCAACGGGGCTCACTG 1619
               ||||||            |||   ::::::   |||:::::::||
     327 GluThrCysCysAspIleAlaThrCysArgPheLysAlaGlySerAsnCy 343

    1620 CAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAG 1669
```

```
     |::::::|||||||||||||:::|||      |||||||||:::   :::
  343 sAlaGluGlyProCysCysGluAsn......CysLeuPheMetSerLysG 358

 1670 GGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGT 1719
             ||||||   :::   :::||||||||||:::   |||||||||
  358 luArgMetCysArgProSerPheGluGluCysAspLeuProGluTyrCys 374

 1720 ACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATA 1769
      :::|||:::|||::::::||||||   |||   :::   |||   |||::
  375 AsnGlySerSerAlaSerCysProGluAsnHisTyrValGlnThrGlyHi 391

 1770 TGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCA 1819
      :   |||:::   ||||||      |||   :::|||   |||   :::
  391 sProCysGlyLeuAsnGlnTrpIleCysIleAspGlyValCysMetSerG 408

 1820 GAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGAC 1869
        |||:::||||||         :::|||   ::::::      :::
  408 lyAspLysGlnCysThrAspThrPheGlyLysGluValGluPheGlyPro 424

 1870 AAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTG 1919
      :::   ||||||:::   ||||||::::::          :::||||||||
  425 SerGluCysTyrSerHisLeuAsnSerLysThrAspValSerGlyAsnCy 441

 1920 CGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCT 1969
      ||||   :::   :::   :::   ||||||:::      ::::::   |
  441 sGlyIleSerAspSerGlyTyrThrGlnCysGluAlaAspAsnLeuGlnC 458

 1970 GTGGATTCTTACTCTGTACCAAT.............CTTACTCGAGCTCCA 2007
      |||||   |||:::|||                  |||   :::   |||
  458 ysGlyLysLeuIleCysLysTyrValGlyLysPheLeuLeuGlnIlePro 474

 2008 CGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCA 2057
      |||         :::   :::::::|||                  |||
  475 ArgAlaThrIleIleTyrAlaAsnIleSerGly.............HisLe 487

 2058 AGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATA 2107
             :::::::::   ::::::   |||:::           |||:
  487 uCysIleAlaValGluPheAlaSerAspHisAla.............AspS 500

 2108 CGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATG 2157
      :::::      ::::::::::|||||||||   ||||||   :::   :::
  500 erGlnLysMetTrpIleLysAspGlyThrSerCysGlySerAsnLysVal 516

 2158 TGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTG 2207
      |||   ::::::::::|||:::::::   :::   |||:::      :::||
  517 CysArgAsnGlnArgCysValSerSerSerTyrLeuGlyTyr...AspCy 532

 2208 TCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTA 2257
      |      |||:::            |||:::   :::|||||||||:::|
  532 sThrThrAspLys.............CysAsnAspArgGlyValCysAsnA 545

 2258 ATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGT 2307
      ||:::      |||   |||:::   ::::::         |||||||||
  545 snLysLysHisCysHisCysSerAlaSerTyrLeuProProAspCysSer 561

 2308 ATCCGG 2313
      ::::::
  562 ValGln 563

seq_name: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-07295-2

seq_documentation_block:
;  Sequence 2, Application PC/TUS9507295
;   GENERAL INFORMATION:
;     APPLICANT:  ALVES, KENNETH
;     APPLICANT:  GUPTA, SUNIL K.
;     APPLICANT:  HOLLIS, GREGORY F.
;     TITLE OF INVENTION:  CONTRACEPTIVE VACCINE
;     NUMBER OF SEQUENCES:  8
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  MARY A. APPOLLINA
;       STREET:  P.O. BOX 2000, 126 E. LINCOLN AVENUE
;       CITY:  RAHWAY
;       STATE:  NJ
;       COUNTRY:  USA
;       ZIP:  07065
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/07295
;       FILING DATE:
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  APPOLLINA, MARY A
;       REGISTRATION NUMBER:  34,087
;       REFERENCE/DOCKET NUMBER:  19244Y
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (908)594-3462
;       TELEFAX:  (908)594-4720
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  651 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
PCT-US95-07295-2


alignment_scores:
            Quality:  711.00              Length:    602
              Ratio:  1.986                Gaps:     21
 Percent Similarity:  59.468   Percent Identity: 28.738

alignment_block:
 US-09-634-252A-2 x PCT-US95-07295-2   ..

 Align seg 1/1  to: PCT-US95-07295-2  from: 1  to: 651

     586 TGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCT 635
         |||:::|||:::|||  |||:::|||          |||   |||   ::
       8 CysHisTyrGlnGlyTyrIleGluGlyTyrProLysSerValValMetVa 24

     636 GTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGT 685
         :|||||||||:::|||||:::|||:::::::::   ::::::      |
      24 lSerThrCysThrGlyLeuArgGlyValLeuGlnPheGluAsnValSerT 41

     686 ATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACAGGTCGA 735
         ||   ||||||||||||              |||:::|||:::|||
      41 yrGlyIleGluProLeu..............GluSerSerValGlyPhe 52

     736 CCACATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAA 785
           |||:::|||   ::::::         :::::::::|||
      53 GluHisValIleTyrGlnVal.........LysHisLysLysAlaAspVa 66

     786 GAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTAC 835
          :::|||      |||:::                          |||   :
      66 lSerLeuTyrAsnGluLysAsp..................IleGluSerA 77

     836 AGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATTT 885
         :.   |||:::   :::   ::::::::::::|||::::::
      77 rgAspLeuSerPheLysLeuGlnSerAlaGluProGlnGln......... 90

     886 GAAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTA 935
         :::      |||||||:::|||:::   ::::::   :::   :::   ||
      91 AspPheAlaLysTyrIleGluMetHisValIleValGluLysGlnLeuTy 107

     936 TAAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCG 985
         |:::         |||:::   :::   :::   :::   :::   ::::
     107 rAsnHisMetGlySerAspThrThrValValAlaGlnLysValPheGlnL 124

     986 TGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTT 1035
         ::::::::|||:::::::::|||:::      ::::::|||      :::
```

```
 124 euIleGlyLeuThrAsnAlaIlePhe...ValSerPheAsnIleThrIle 139

1036 GTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCAC 1085
     :::|||   ::::::|||   |||   :::::::::::::|||      ||
 140 IleLeuSerSerLeuGluLeuTrpIleAspGluAsnLysIleAlaThrTh 156

1086 CACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGC. 1134
     |   :::      ::::::||||||   |||   :::::::::
 156 rGlyGluAlaAsnGluLeuLeuHisThrPheLeuArgTrpLysThrSerT 173

1135 ........ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGG... 1173
             ::::::   |||      |||      |||:::   |||
 173 yrLeuValLeuArgProHisAspValAlaPheLeuLeuValTyrArgGlu 189

1174 ..................GTGACATTTCACTATAAGAGAAGCAGTCTGAG 1205
                       :::||||||:::   |||      :::   ::
 190 LysSerAsnTyrValGlyAlaThrPheGlnGlyLysMetCysAspAlaAs 206

1206 TTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGT 1255
     :|||   |||||||||      :::   |||   ::::::::::::
 206 nTyrAlaGlyGlyValValLeuHisProArgThrIleSerLeuGlu.... 221

1256 ATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAA 1305
                  ::::::|||   :::|||:::|||   |||:::
 222 ..............SerLeuAlaValIleLeuAlaGlnLeuLeuSerLeu 233

1306 AACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATGTGACTG 1355
     ::::::||||||   ::::::      :::         ||||||:::||
 234 SerMetGlyIleThrTyrAspAspIleAsn.........LysCysGlnCy 247

1356 CACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCATT 1405
     |      |||   :::   |||||||||:::      :::::      |
 247 s......SerGlyAlaValCysIleMetAsnProGluAlaIleHisPheS 262

1406 CT......CGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTT 1449
     ||      :::   ||||||:::||||||:::   ::::::      |||
 262 erGlyValLysIlePheSerAsnCysSerPheGluAspPheAlaHisPhe 278

1450 TTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATT 1499
     :::::::::      :::   ||||||   |||:::|||      ::
 279 IleSerLysGlnLysSerGlnCysLeuHisAsnGlnProArg......Le 293

1500 TGAGCCC..............ACGGAATGTGGAAATGGATACGTGGAAG 1534
     ::::|||                ::::   ||||||||||:::   :::||||
 293 uAspProPhePheLysGlnGlnAlaValCysGlyAsnAlaLysLeuGluA 310

1535 CTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTATGGATTA... 1581
     ||||||||||||||||||||||||   :::   :::|||      :::
 310 laGlyGluGluCysAspCysGlyThrGluGlnAspCysAlaLeuIleGly 326

1582 ......TGCTGT......AAGAAATGTTCCCTCTCCAACGGGGCTCACTG 1619
           ||||||          |||   ::::::   |||:::::::||
 327 GluThrCysCysAspIleAlaThrCysArgPheLysAlaGlySerAsnCy 343

1620 CAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAG 1669
     |:::::::||||||||||||:::|||    . |||||||||:::   :::
 343 sAlaGluGlyProCysCysGluAsn......CysLeuPheMetSerLysG 358

1670 GGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGT 1719
            ||||||   :::   :::|||||||||:::   |||||||||
 358 luArgMetCysArgProSerPheGluGluCysAspLeuProGluTyrCys 374

1720 ACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATA 1769
     :::|||:::|||::::::||||||   |||   :::   |||   |||::
 375 AsnGlySerSerAlaSerCysProGluAsnHisTyrValGlnThrGlyHi 391

1770 TGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCA 1819
     :   |||:::   ||||||      |||   :::|||   |||   :::
 391 sProCysGlyLeuAsnGlnTrpIleCysIleAspGlyValCysMetSerG 408

1820 GAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGAC 1869
       |||:::||||||         :::|||   ::::::      :::
 408 lyAspLysGlnCysThrAspThrPheGlyLysGluValGluPheGlyPro 424
```

```
1870 AAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTG 1919
     :::   ||||||:::   ||||||::::::         :::||||||||
 425 SerGluCysTyrSerHisLeuAsnSerLysThrAspValSerGlyAsnCy 441

1920 CGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCT 1969
     ||||   :::   :::   :::   ||||||:::      ::::::   |
 441 sGlyIleSerAspSerGlyTyrThrGlnCysGluAlaAspAsnLeuGlnC 458

1970 GTGGATTCTTACTCTGTACCAAT.............CTTACTCGAGCTCCA 2007
     |||||   |||:::|||                   |||   :::   |||
 458 ysGlyLysLeuIleCysLysTyrValGlyLysPheLeuLeuGlnIlePro 474

2008 CGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCA 2057
     |||          :::   ::::::|||                   |||
 475 ArgAlaThrIleIleTyrAlaAsnIleSerGly.............HisLe 487

2058 AGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATA 2107
            :::::::::   ::::::   |||:::            |||:
 487 uCysIleAlaValGluPheAlaSerAspHisAla.............AspS 500

2108 CGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATG 2157
     :::::      ::::::::::|||||||||   ||||||   :::   :::
 500 erGlnLysMetTrpIleLysAspGlyThrSerCysGlySerAsnLysVal 516

2158 TGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTG 2207
     |||   ::::::::::|||::::::   :::   |||:::      :::||
 517 CysArgAsnGlnArgCysValSerSerSerTyrLeuGlyTyr...AspCy 532

2208 TCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTA 2257
     |      |||:::          |||:::   :::|||||||||:::|
 532 sThrThrAspLys............CysAsnAspArgGlyValCysAsnA 545

2258 ATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGT 2307
     ||:::      |||   |||:::   ::::::         |||||||||
 545 snLysLysHisCysHisCysSerAlaSerTyrLeuProProAspCysSer 561

2308 ATCCGG 2313
     ::::::
 562 ValGln 563
```

```
seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-765-243-6

seq_documentation_block:
; Sequence 6, Application US/08765243
; Patent No. 5935578
;  GENERAL INFORMATION:
;    APPLICANT:  ALVES, KENNETH
;    APPLICANT:  GUPTA, SUNIL K.
;    APPLICANT:  HOLLIS, GREGORY F.
;    TITLE OF INVENTION:  CONTRACEPTIVE VACCINE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MARY A. APPOLLINA
;      STREET:  P.O. BOX 2000, 126 E. LINCOLN AVENUE
;      CITY:  RAHWAY
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07065
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/765,243
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  APPOLLINA, MARY A
;      REGISTRATION NUMBER:  34,087
;      REFERENCE/DOCKET NUMBER:  19244Y
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  (908)594-3462
;      TELEFAX:  (908)594-4720
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  735 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-765-243-6


alignment_scores:
           Quality:  691.00              Length:    675
             Ratio:  1.799                Gaps:     26
rcent Similarity:  56.889    Percent Identity: 28.741

[redacted]ment_block:
09-634-252A-2 x US-08-765-243-6   ..

 Align seg 1/1  to: US-08-765-243-6  from: 1  to: 735

     448 CAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCT 497
         |||:::::::::::::::::::      |||:::   :::   ||||||||
      50 GlnValThrTyrAsnLeuLysIleGluGlyLysThrTyrThrLeuAspLe 66

     498 CATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTACG 547
         |   :::::::::   :::|||      ::::::             |||:
      66 u...MetGlnLysProPheLeuProProAsnPheArgValTyrSerTyrA 82

     548 AAAATGGG.........AAACCACAGTACTCTAAGGGTGGAGAGCACTGT 588
         ::|||:::          :::         :::|||      :::   |||
      82 spAsnAlaGlyIleMetArgSerLeuGluGlnLysPheGlnAsnIleCys 98

     589 TACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTC 638
         |||:::::::|||   |||:::|||      :::|||   |||   :::||
      99 TyrPheGlnGlyTyrIleGluGlyTyrProAsnSerMetValIleValSe 115

     639 AACCTGCAATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATA 688
         |||||||:::|||||||:::|||:::::::::       :::       |||
     115 rThrCysThrGlyLeuArgGlyPheLeuGlnPheGlyAsnValSerTyrG 132

     689 TGATAGAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCA 738
           ||||||||||||               |||:::|||:::|||
     132 lyIleGluProLeu..............GluSerSerSerGlyPheGlu 143

     739 CATATAATC..............CAGAAAACCTTGGCAGGACAGTATTC 773
         |||:::|||            :::|||      |||      |||::
     144 HisValIleTyrGlnValGluProGluLysGlyGlyAlaLeuLeuTyrAl 160

     774 TAAGCAAATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTC 823
         :::::::      :::|||          |||   :::|||:::
     160 aGluLysAspIleAspLeu.........ArgAspSerGlnTyr........ 171

     824 TCTCTGAATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCA 873
                                :::   |||::::::::::|||:::
     172 ........................LysIleArgSerIleLysProGln 179

     874 CGTGGTATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAATGA 923
         |||   :::                |||||||||:::   ||||||   ::
     180 ArgIleVal.........SerHisTyrLeuGluIleHisIleValAsnGl 193

     924 TCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAACAACT 973
         :   :::   ::::::          ::::::   |||   :::::::::
     193 uLysGlnMetPheGluHisIleGlyAlaAspThrAlaIleValThrGlnL 210

     974 TTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTC 1023
                 ::::::::::::|||:::::::::|||:::          :::
     210 ysIlePheGlnLeuIleGlyLeuAlaAsnAlaIlePhe...AlaProPhe 225

    1024 AACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCA 1073
         |||      |||:::|||   :::::::|||   |||   ::::::::::
     226 AsnLeuThrValIleLeuSerSerLeuGluPheTrpMetAspGluAsnLy 242

    1074 GATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAAT 1123
         :|||      |||   :::      ::::::|||:::::::|||   |||:
     242 sIleLeuThrThrGlyAspAlaAsnLysLeuLeuTyrArgPheLeuLysT 259

    1124 ACCGGCAGCGC.........ATTAAGCAGCATGCTGATGCTGTGCACCTC 1164
         :::::|||            ::::::   |||      |||      |||
     259 rpLysGlnSerTyrLeuValLeuArgProHisAspMetAlaPheLeuLeu 275

    1165 ATCTCGCGGGTGACATTTCACTAT.....................AAGAG 1193
         :::   |||   |||      |||                      |||
     276 ValTyrArgAsnThrThrAspTyrValGlyAlaThrTyrGlnGlyLysMe 292

    1194 AAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTG 1243
            :::   :::|||   |||||||||:::   :::   ::::::|||
     292 tCysAspLysAsnTyrAlaGlyGlyValAlaLeuHisProLysAlaValT 309

    1244 GTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAG 1293
           ::::::                    :::::::|||   :::|||   |||
     309 hrLeuGlu.................SerLeuAlaIleIleLeuValGln 319

    1294 AGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCC 1343
            |||:::   ::::::|||:::   ::::::      :::
     320 LeuLeuSerLeuSerMetGlyLeuAlaTyrAspAspValAsn........ 333

    1344 AAAATGTGACTGCACAGAATCCTGGGGTGGC.........TGCATCATG. 1383
          ||||||:::|||               |||          |||:::|||
     334 .LysCysGlnCys..............GlyValProValCysValMetA 345

    1384 .....GAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGC 1428
          |||             |||      |||   ||||||:::||||||
     345 snProGluAlaProHisSerSerGlyValArgAlaPheSerAsnCysSer 361

    1429 ATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTT 1478
         :::   ::::::      |||:::      ::::::   ||||||
     362 MetGluAspPheSerLysPheIleThrSerGlnSerSerHisCysLeuGl 378

    1479 CAACAGGCCAACA......AAGCTATTTGAGCCCACGGAATGTGGAAATG 1522
          |||:::||||||            .::::::   :::   ||||||||||
     378 nAsnGlnProThrLeuGlnProSerTyrLysMetAlaValCysGlyAsnG 395

    1523 GATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCAT......GTG 1566
         ||   ||||||      |||  ||||||||||||             :::
     395 lyGluValGluGluAspGluIleCysAspCysGlyLysLysGlyCysAla 411

    1567 GAATGCTATGGATTATGCTGTAAG......AAATGTTCCCTCTCCAACGG 1610
         |||            ||||||:::       |||:::||||||:::||
     412 GluMetProProProCysCysAsnProAspThrCysLysLeuSerAspGl 428

    1611 GGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTC 1660
         |:::::|||||:::|||   |||||||||      ||||||   ::::
     428 ySerGluCysSerSerGlyIleCysCysAsn......SerCysLysLeuL 443

    1661 AGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACT 1710
         ::   :::|||      ||||||   |||   :::|||||||||:::|||
     443 ysArgLysGlyGluValCysArgLeuAlaGlnAspGluCysAspValThr 459

    1711 GAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCA 1760
         |||||||||:::|||   |||      |||      ::::::      ||
     460 GluTyrCysAsnGlyThrSerGluValCys...GluAspPhePheValGl 475

    1761 AGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGT 1810
         |:::|||:::   |||:::::::::::      |||   ||||||   |
     475 nAsnGlyHisProCysAspAsnArgLysTrpIleCysIleAsnGlyThrC 492

    1811 GCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCA 1860
         ||::::::   ::::::|||||||||   :::::::|||      |||:::
     492 ysGlnSerGlyGluGlnGlnCysGlnAspLeuPheGlyIleAspAlaGly 508

    1861 GGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAA 1910
            :::::::::   |||:::   :::||||||:::::::::      ::
     509 PheGlySerSerGluCysPheTrpGluLeuAsnSerLysSerAspIleSe 525

    1911 GGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATG 1960
```

```
      :|||:::||||||   ::::::      :::   :::|||      :::|
   525 rGlySerCysGlyIleSerAlaGlyGlyTyrLysGluCysProProAsnA 542

  1961 ATGTGTTCTGTGGATTCTTACTCTGT.....................ACC 1989
       ||   :::|||||||   :::::|||
   542 spArgMetCysGlyLysIleIleCysLysTyrGlnSerGluAsnIleLeu 558

  1990 AATCTTACTCGAGCTCCACGTATT...GGTCAACTTCAGGGTGAGATCAT 2036
       :::|||      |||      |||   :::::::::::|||::::::
   559 LysLeuArgSerAlaThrValIleTyrAlaAsnIleSerGlyHisValCy 575

  2037 TCCAACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCC 2086
           :::      |||   ||||||:::
   575 sValSerLeuGluTyrProGlnGlyHis..................... 584

  2087 ATGTAGTTTTAGATGATGATACGGATGTGGGCTATGTAGAAGATGGAACG 2136
                    ::::::::::::      :::|||:::|||||||||
   585 .............AsnGluSerGlnLysMetTrpValArgAspGlyThr 596

  2137 CCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAATTCA 2186
       ||||||   :::   :::|||   ::::::||||||:::
   597 ValCysGlySerAsnLysValCysGlnAsnGlnLysCysValAlaAspTh 613

  2187 AGCCCTAAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTT 2236
         |||:::      :::|||   |||::::::          |||:
   613 rPheLeuGlyTyr...AspCysAsnLeuGluLys.....CysA 625

  2237 CGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACC 2286
       ::   ||||||||||||:::|||:::   :::|||   ||||||   |||
   625 snHisHisGlyValCysAsnAsnLysLysAsnCysHisCysAspProThr 641

  2287 TGGGCAGGGACAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCC 2336
       :::         ||||||:::
   642 TyrLeuProProAspCysLys.....................Ar 649

  2337 CCCCAAGGATGAAGGACCCAAGGGT 2361
         ||||||:::   |||   |||
   649 gMetLysAspSerTyrProGlyGly 657

seq_name: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-07295-6

seq_documentation_block:
; Sequence 6, Application PC/TUS9507295
;  GENERAL INFORMATION:
;    APPLICANT:  ALVES, KENNETH
;    APPLICANT:  GUPTA, SUNIL K.
;    APPLICANT:  HOLLIS, GREGORY F.
;    TITLE OF INVENTION:  CONTRACEPTIVE VACCINE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MARY A. APPOLLINA
;      STREET:  P.O. BOX 2000, 126 E. LINCOLN AVENUE
;      CITY:  RAHWAY
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07065
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/07295
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  APPOLLINA, MARY A
;      REGISTRATION NUMBER:  34,087
;      REFERENCE/DOCKET NUMBER:  19244Y
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)594-3462
;      TELEFAX:  (908)594-4720
;  INFORMATION FOR SEQ ID NO:  6:

;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  735 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US95-07295-6


alignment_scores:
            Quality:  691.00              Length:    675
              Ratio:  1.799                 Gaps:     26
 Percent Similarity:  56.889   Percent Identity: 28.741

alignment_block:
 US-09-634-252A-2 x PCT-US95-07295-6   ..

 Align seg 1/1  to: PCT-US95-07295-6  from: 1  to: 735

    448 CAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCT 497
        |||:::::::::::::::::::      |||:::   :::   ||||||||
     50 GlnValThrTyrAsnLeuLysIleGluGlyLysThrTyrThrLeuAspLe 66

    498 CATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTACG 547
        |   :::::::::   :::|||      ::::::            |||:
     66 u...MetGlnLysProPheLeuProProAsnPheArgValTyrSerTyrA 82

    548 AAAATGGG.........AAACCACAGTACTCTAAGGGTGGAGAGCACTGT 588
        ::|||:::         :::        :::|||      :::   |||
     82 spAsnAlaGlyIleMetArgSerLeuGluGlnLysPheGlnAsnIleCys 98

    589 TACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTC 638
        |||:::::|||   |||::|||      :::|||   |||   :::||
     99 TyrPheGlnGlyTyrIleGluGlyTyrProAsnSerMetValIleValSe 115

    639 AACCTGCAATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATA 688
        |||||||:::|||||||:::|||:::::::::      :::      |||
    115 rThrCysThrGlyLeuArgGlyPheLeuGlnPheGlyAsnValSerTyrG 132

    689 TGATAGAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCA 738
          ||||||||||||               |||:::|||:::|||
    132 lyIleGluProLeu..............GluSerSerSerGlyPheGlu 143

    739 CATATAATC..............CAGAAAACCTTGGCAGGACAGTATTC 773
        |||:::|||              :::|||      |||      |||::
    144 HisValIleTyrGlnValGluProGluLysGlyGlyAlaLeuLeuTyrAl 160

    774 TAAGCAAATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTC 823
        :::::::      :::|||          |||   :::|||:::
    160 aGluLysAspIleAspLeu.........ArgAspSerGlnTyr....... 171

    824 TCTCTGAATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCA 873
                                     :::   |||::::::::::|||:::
    172 .........................LysIleArgSerIleLysProGln 179

    874 CGTGGTATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAATGA 923
        |||   :::               ||||||||||:::   ||||||   ::
    180 ArgIleVal.........SerHisTyrLeuGluIleHisIleValValGl 193

    924 TCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAACAACT 973
        :   :::   ::::::         ::::::   |||   :::::::::
    193 uLysGlnMetPheGluHisIleGlyAlaAspThrAlaIleValThrGlnL 210

    974 TTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTC 1023
              ::::::::::::|||:::::::::|||:::         :::
    210 ysIlePheGlnLeuIleGlyLeuAlaAsnAlaIlePhe...AlaProPhe 225

   1024 AACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCA 1073
        |||      |||:::|||   :::::|||   |||   ::::::::::
    226 AsnLeuThrValIleLeuSerSerLeuGluPheTrpMetAspGluAsnLy 242

   1074 GATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAAT 1123
        :|||      |||   :::      ::::::|||::::::|||   |||:
    242 sIleLeuThrThrGlyAspAlaAsnLysLeuLeuTyrArgPheLeuLysT 259
```

```
1124 ACCGGCAGCGC.........ATTAAGCAGCATGCTGATGCTGTGCACCTC 1164
     :::::|||            ::::::   |||      |||      |||
 259 rpLysGlnSerTyrLeuValLeuArgProHisAspMetAlaPheLeuLeu 275

1165 ATCTCGCGGGTGACATTTCACTAT.....................AACAG 1193
     :::   |||   |||      |||                      |||
 276 ValTyrArgAsnThrThrAspTyrValGlyAlaThrTyrGlnGlyLysMe 292

1194 AAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTG 1243
       :::   :::|||  ||||||||||:::   :::   ::::::|||
 292 tCysAspLysAsnTyrAlaGlyGlyValAlaLeuHisProLysAlaValT 309

1244 GTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAG 1293
       ::::::                  ::::::|||   :::|||   |||
 309 hrLeuGlu.................SerLeuAlaIleIleLeuValGln 319

1294 AGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCC 1343
       |||:::   :::::::|||:::   ::::::      :::
 320 LeuLeuSerLeuSerMetGlyLeuAlaTyrAspAspValAsn.....   333

1344 AAAATGTGACTGCACAGAATCCTGGGGTGGC.........TGCATCATG. 1383
      ||||||:::|||                |||        |||:::|||
 334 .LysCysGlnCys...............GlyValProValCysValMetA 345

1384 .....GAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGC 1428
          |||            |||      |||   ||||||:::||||||
 345 snProGluAlaProHisSerSerGlyValArgAlaPheSerAsnCysSer 361

1429 ATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTT 1478
     :::   ::::::      |||:::          ::::::   ||||||
 362 MetGluAspPheSerLysPheIleThrSerGlnSerSerHisCysLeuGl 378

1479 CAACAGGCCAACA......AAGCTATTTGAGCCCACGGAATGTGGAAATG 1522
      |||:::||||||                ::::::   :::   ||||||||||
 378 nAsnGlnProThrLeuGlnProSerTyrLysMetAlaValCysGlyAsnG 395

1523 GATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCAT......GTG 1566
     ||   ||||||      |||   ||||||||||||             :::
 395 lyGluValGluGluAspGluIleCysAspCysGlyLysLysGlyCysAla 411

1567 GAATGCTATGGATTATGCTGTAAG......AAATGTTCCCTCTCCAACGG 1610
     |||            ||||||:::          |||:::||||||:::||
 412 GluMetProProProCysCysAsnProAspThrCysLysLeuSerAspGl 428

1611 GGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTC 1660
     |:::::|||||||:::|||   |||||||||      ||||||   ::::
 428 ySerGluCysSerSerGlyIleCysCysAsn......SerCysLysLeuL 443

1661 AGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACT 1710
     ::   :::|||      ||||||   |||   :::||||||||||:::|||
 443 ysArgLysGlyGluValCysArgLeuAlaGlnAspGluCysAspValThr 459

1711 GAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCA 1760
     |||||||||:::|||   |||      |||      ::::::      ||
 460 GluTyrCysAsnGlyThrSerGluValCys...GluAspPhePheValGl 475

1761 AGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGT 1810
     |:::|||:::   |||:::::::::::::      |||   ||||||   |
 475 nAsnGlyHisProCysAspAsnArgLysTrpIleCysIleAsnGlyThrC 492

1811 GCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCA 1860
     ||::::::   ::::::|||||||||   ::::::|||      |||:::
 492 ysGlnSerGlyGluGlnGlnCysGlnAspLeuPheGlyIleAspAlaGly 508

1861 GGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGACAA 1910
        :::::::::   |||:::   :::||||||:::::::::      ::
 509 PheGlySerSerGluCysPheTrpGluLeuAsnSerLysSerAspIleSe 525

1911 GGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATG 1960
     :|||:::||||||   ::::::      :::   :::|||      :::|
 525 rGlySerCysGlyIleSerAlaGlyGlyTyrLysGluCysProProAsnA 542

1961 ATGTGTTCTGTGGATTCTTACTCTGT.....................ACC 1989
       ||   :::||||||   ::::::|||
  542 spArgMetCysGlyLysIleIleCysLysTyrGlnSerGluAsnIleLeu 558

1990 AATCTTACTCGAGCTCCACGTATT...GGTCAACTTCAGGGTGAGATCAT 2036
       :::|||      |||      |||   ::::::::::::|||::::::
  559 LysLeuArgSerAlaThrValIleTyrAlaAsnIleSerGlyHisValCy 575

2037 TCCAACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCC 2086
           :::      |||   |||||||:::
  575 sValSerLeuGluTyrProGlnGlyHis..................... 584

2087 ATGTAGTTTTAGATGATGATACGGATGTGGGCTATGTAGAAGATGGAACG 2136
                    ::::::::::::      :::|||:::||||||||||
  585 .............AsnGluSerGlnLysMetTrpValArgAspGlyThr 596

2137 CCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAATTCA 2186
        ||||||   :::   :::|||   ::::::||||||:::
  597 ValCysGlySerAsnLysValCysGlnAsnGlnLysCysValAlaAspTh 613

2187 AGCCCTAAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTT 2236
          |||:::      :::|||   |||::::::              |||:
  613 rPheLeuGlyTyr...AspCysAsnLeuGluLys.............CysA 625

2237 CGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACC 2286
       ::   ||||||||||||:::|||:::   :::|||   ||||||   |||
  625 snHisHisGlyValCysAsnAsnLysLysAsnCysHisCysAspProThr 641

2287 TGGGCAGGGACAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCC 2336
       :::         ||||||:::
  642 TyrLeuProProAspCysLys..........................Ar 649

2337 CCCCAAGGATGAAGGACCCAAGGGT 2361
           |||||||:::   |||   |||
  649 gMetLysAspSerTyrProGlyGly 657
```

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-836-442-3

```
seq_documentation_block:
; Sequence 3, Application US/08836442
; Patent No. 5990293
;  GENERAL INFORMATION:
;    APPLICANT:  DOCHERTY, Andrew, J.P.
;    APPLICANT:  SLOCOMBE, Patrick, M.
;    TITLE OF INVENTION:  A HUMAN METALLOPROTEINASE
;    TITLE OF INVENTION:  VARIANTS THEREOF AND DNA SEQUENCES CODING THEREFOR
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,442
;      FILING DATE:  01-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/GB96/02181
;      FILING DATE:  13-MAR-1997
;      APPLICATION NUMBER:  GB 9612150.4
;      FILING DATE:  11-JUN-1996
;      APPLICATION NUMBER:  GB 9526229.1
;      FILING DATE:  21-DEC-1995
;      APPLICATION NUMBER:  GB 9521498.7
;      FILING DATE:  20-OCT-1995
;      APPLICATION NUMBER:  GB 95521495.3
```

```
;      FILING DATE:  20-OCT-1995
;      APPLICATION NUMBER:  GB 9518023.8
;      FILING DATE:  05-SEP-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Resnick, David S
;      REGISTRATION NUMBER:  34,235
;      REFERENCE/DOCKET NUMBER:  47425
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  529 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  No. 5990293e
US-08-836-442-3

[redacted]nment_scores:
            Quality:  577.50              Length:    525
              Ratio:  1.887               Gaps:       20
 Percent Similarity:  58.286   Percent Identity: 30.857

alignment_block:
 US-09-634-252A-2 x US-08-836-442-3   ..

 Align seg 1/1  to: US-08-836-442-3  from: 1  to: 529

     328 GAAATCACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGA 377
         |||::::::   |||   ||||||
      29 GluValValTyrProIleArgLeu..........................  36

     378 AAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATA 427
             |||   |||       :::::::|||      :::||||||::::::
      37 .HisProLeuHisLysArgGluAlaLysGluProGluGlnGlnGluGlnP  53

     428 ATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGC 477
           :::           ::::::   ::::::::::::         |||
      53 heGlu............ThrGluLeuLysTyrLysMetThrIleAsnGly  65

     478 TCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGA 527
         :::       :::|||   |||      |||:::::::||||||:::
      66 LysIleAlaValLeuTyrLeuLysLysAsnLysAsnLeuLeuAlaProGl  82

     528 TTATGTGGAGATTCACTACGAAAAT...GGGAAA..............C 559
          |||:::|||   :::|||::::::   ||||||             |
      82 yTyrThrGluThrTyrTyrAsnSerThrGlyLysGluIleThrThrSerP  99

     560 CACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGA 609
         |||||               :::   ||||||||||:::|||   |||
      99 roGlnIleMet.........AspAspCysTyrTyrGlnGlyHisIleLeu 112

     610 GGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGG 659
         :::   |||   |||   :::::::::|||||||||:::||||||:::||
     113 AsnGluLysValSerAspAlaSerIleSerThrCysArgGlyLeuArgGl 129

     660 CATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGG 709
         |   |||::::::              |||:::||||||||||||:::   :
     129 yTyrPheSerGlnGlyAspGlnArgTyrPheIleGluProLeuSerProI 146

     710 TTCATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAACCTTG 759
         ::|||   :::         |||:::   |||   :::   |||
     146 leHisArgAsp.........GlyGlnGluHisAlaLeuPheLys...... 157

     760 GCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAGGT.. 807
                 |||:::   :::   ||||||                |||
     158 .........TyrAsnProAspGluLysAsnTyrAspSerThrCysGlyMe 171

     808 .GACCAGTGGCCCTTTCTCTCTGAATTACAGTGG.........TTGAAAA 847
          |||         :::      :::|||||||             |||

     171 tAspGlyValLeuTrpAlaHisAspLeuGlnGlnAsnIleAlaLeuProA 188

     848 GAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAA 897
             |||   :::          :::|||   :::   |||:::   |||
     188 laThrLysLeuValLysLeuLysAspArgLysValGlnGluHisGluLys 204

     898 TATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCG 947
         |||:::|||      :::|||   |||:::      :::|||:::::::
     205 TyrIleGluTyrTyrLeuValLeuAspAsnGlyGluPheLysArgTyrAs 221

     948 CTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTG 997
         ::::::::::   :::   ::::::   :::   :::::::::|||   |
     221 nGluAsnGlnAspGluIleArgLysArgValPheGluMetAlaAsnTyrV 238

     998 TGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCT 1047
         ||:::   :::||||||   :::||||||||||:::|||:::|||||:::
     238 alAsnMetLeuTyrLys...LysLeuAsnThrHisValAlaLeuValGly 253

    1048 GTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGT 1097
         :::|||   ||||||:::||||||:::|||   ||||||   |||
     254 MetGluIleTrpThrAspLysAspLysIleLysIleThrProAsnAlaSe 270

    1098 GCAGATGCTCCATGAGTTCTCAAAATACCGG.....CAGCGCA 1135
                |||::::::||||||||:::|||          :::|||
     270 rPheThrLeuGluAsnPheSerLysTrpArgGlySerValLeuSerArgA 287

    1136 TTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCAC 1185
           |||:::|||   |||   ::::::||||||:::   :::   :::
     287 rgLysArgHis...AspIleAlaGlnLeuIleThrAlaThrGluLeuAla 302

    1186 TATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCT...CGCAC 1232
                      :::|||:::::::::::::   :::||||||      ::
     303 GlyThrThrValGlyLeuAlaPheMetSerThrMetCysSerProTyrSe 319

    1233 AAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACAAG 1282
         :   ||||||      ::::::            :::   ||||||   :
     319 rValGlyValValGlnAspHisSerAspAsnLeuLeuArgValAlaGlyT 336

    1283 TATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCT 1332
         :::::::::::::::::::::::|||:::|||:::      :::
     336 hrMetAlaHisGluMetGlyHisAsnPheGlyMetPheHisAspAspTyr 352

    1333 AGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCAT 1382
         |||             |||   |||   ::::::          |||:::||
     353 Ser............CysLysCysProSerThrIle......CysValMe 363

    1383 GGAGGAAACA...GGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCA 1429
         |:::::::::   :::                  ||||||:::||||||
     363 tAspLysAlaLeuSerPheTyrIleProThrAspPheSerSerCysSerA 380

    1430 TTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTC 1479
           |||:::|||      |||::::::         :::   |||||||||
     380 rgLeuSerTyrAspLysPhePheGluAspLysLeuSerAsnCysLeuPhe 396

    1480 AACAGG......CCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGG 1523
         |||          ||||||   :::::::::          |||||||||
     397 AsnAlaProLeuProThrAspIleIleSerThrProIleCysGlyAsnGl 413

    1524 ATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCT 1573
             ||||||   ||||||:::||||||||||||
     413 nLeuValGluMetGlyGluAspCysAspCysGly.................. 424

    1574 ATGGATTATGCTGTAAGAAATGTTCCCTCTCCAACGGGGCTCACTGCAGC 1623
                       :::   |||::::::::::         :::|||:::
     425 ........ThrSerGluThrCysLysIleLysAlaThrPheGlnCysAla 438

    1624 GACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGTA 1673
           |||   ||||||:::      :::|||   |||:::      |||
     439 LeuGlyGluCysCysGlu......LysCysGlnPheLysLysAlaGlyMe 453

    1674 TGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTACTG 1723
         ||||||   |||   :::||||||||||:::   |||   |||:::|
     453 tValCysArgProAlaLysAspGluCysAspLeuProGluMetCysAsnG 470
```

```
    1724 GAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATGCA 1773
         ||   ||||||:::|||||||   :::      :::   :::|||:::
     470 lyLysSerGlyAsnCysProAspAspArgPheGlnValAsnGlyPhePro 486

    1774 TGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGAGA 1823
         |||:::   :::::::|||:::|||      |||   |||   |||   ::
     487 CysHis...GlyLysGlyHisCysLeuMetGlyThrCysProThrLeuGl 502

    1824 CAACCAGTGTCAGTACATCTGGGGA 1848
         ::::||||||      :::||||||
     502 nGluGlnCysThrGluLeuTrpGly 510

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-264-101-4

;  documentation_block:
;  equence 4, Application US/08264101
;  atent No. 5693496
;  GENERAL INFORMATION:
;    APPLICANT:  ALVES, KENNETH
;    APPLICANT:  GUPTA, SUNIL K.
;    APPLICANT:  HOLLIS, GREGORY F.
;    TITLE OF INVENTION:  CONTRACEPTIVE VACCINE
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MARY A. APPOLLINA
;      STREET:  P.O. BOX 2000, 126 E. LINCOLN AVENUE
;      CITY:  RAHWAY
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07065
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/264,101
;      FILING DATE:
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  APPOLLINA, MARY A.
;      REGISTRATION NUMBER:  34,087
;      REFERENCE/DOCKET NUMBER:  19244
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)594-3462
;      TELEFAX:  (908)594-4720
;   NFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  457 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-264-101-4


alignment_scores:
            Quality:  517.00              Length:    414
              Ratio:  2.163                 Gaps:     16
 Percent Similarity:  57.729   Percent Identity: 31.159

alignment_block:
 US-09-634-252A-2 x US-08-264-101-4   ..

 Align seg 1/1  to: US-08-264-101-4  from: 1  to: 457

    1177 ACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTC 1226
         |||::::::   |||      :::   :::|||   |||||||||:::
       9 ThrTyrGlnGlyLysMetCysAspLysAsnTyrAlaGlyGlyValAlaLe 25

    1227 TCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGG 1276
          :::   ::::::|||   ::::::                   ::::::|

      25 uHisProLysAlaValThrLeuGlu.................SerLeuA 36

    1277 CACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAA 1326
         ||   :::|||   |||   |||:::   ::::::|||:::   ::::::
      36 laIleIleLeuValGlnLeuLeuSerLeuSerMetGlyLeuAlaTyrAsp 52

    1327 CCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGC.. 1374
               :::         ||||||:::|||               |||
      53 AspValAsn.........LysCysGlnCys..............GlyVa 61

    1375 .......TGCATCATG......GAGGAAACAGGGGTGTCCCATTCTCGAA 1411
               |||:::|||      |||             |||      |||
      61 lProValCysValMetAsnProGluAlaProHisSerSerGlyValArgA 78

    1412 AATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGA 1461
           ||||||:::||||||:::   ::::::      |||:::
      78 laPheSerAsnCysSerMetGluAspPheSerLysPheIleThrSerGln 94

    1462 GGTGGAGCCTGCCTTTTCAACAGGCCAACA......AAGCTATTTGAGCC 1505
         ::::::   ||||||   |||:::||||||           ::::::
      95 SerSerHisCysLeuGlnAsnGlnProThrLeuGlnProSerTyrLysMe 111

    1506 CACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTG 1555
          :::   ||||||||||||   ||||||      |||   ||||||||||
     111 tAlaValCysGlyAsnGlyGluValGluGluAspGluIleCysAspCysG 128

    1556 GTTTTCAT......GTGGAATGCTATGGATTATGCTGTAAG......AAA 1593
         ||           :::|||              ||||||:::
     128 lyLysLysGlyCysAlaGluMetProProProCysCysAsnProAspThr 144

    1594 TGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAA 1643
         |||:::|||||||:::|||:::::::||||||:::|||   |||||||||
     145 CysLysLeuSerAspGlySerGluCysSerSerGlyIleCysCysAsn.. 160

    1644 TACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGA 1693
         ||||||   ::::::   :::|||      ||||||   |||   :
     161 ....SerCysLysLeuLysArgLysGlyGluValCysArgLeuAlaGlnA 176

    1694 ACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCA 1743
         ::|||||||||:::|||||||||||||:::|||   |||      |||
     176 spGluCysAspValThrGluTyrCysAsnGlyThrSerGluValCys... 191

    1744 CCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCG 1793
         ::::::      |||:::|||:::   |||:::::::::::::
     192 GluAspPhePheValGlnAsnGlyHisProCysAspAsnArgLysTrpIl 208

    1794 CTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCT 1843
          |||   ||||||   |||:::::::   ::::::||||||||   ::::
     208 eCysIleAsnGlyThrCysGlnSerGlyGluGlnGlnCysGlnAspLeuP 225

    1844 GGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAAT 1893
         ::|||      |||:::   :::::::::   |||:::   :::||||||
     225 heGlyIleAspAlaGlyPheGlySerSerGluCysPheTrpGluLeuAsn 241

    1894 ACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGAT 1943
         :::::::::      :::|||:::||||||   ::::::      :::
     242 SerLysSerAspIleSerGlySerCysGlyIleSerAlaGlyGlyTyrLy 258

    1944 TCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGT....... 1986
          :::|||      :::|||   :::||||||   ::::::|||
     258 sGluCysProProAsnAspArgMetCysGlyLysIleIleCysLysTyrG 275

    1987 ..............ACCAATCTTACTCGAGCTCCACGTATT...GGTCAA 2019
                        :::|||      |||      |||   ::::::
     275 lnSerGluAsnIleLeuLysLeuArgSerAlaThrValIleTyrAlaAsn 291

    2020 CTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGAT 2069
         :::::::|||::::::      :::      |||   |||||||:::
     292 IleSerGlyHisValCysValSerLeuGluTyrProGlnGlyHis..... 306

    2070 TGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCT 2119
                                          ::::::::::::      :
     307 ............................AsnGluSerGlnLysMetT 313
```

```
   2120 ATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGG 2169
        ::|||:::|||||||||   ||||||   :::   :::|||   ::::::
    313 rpValArgAspGlyThrValCysGlySerAsnLysValCysGlnAsnGln 329

   2170 AAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTC 2219
        ||||||:::           |||:::      :::|||   |||:::::
    330 LysCysValAlaAspThrPheLeuGlyTyr...AspCysAsnLeuGluLy 345

   2220 CAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCT 2269
        :           |||:::   |||||||||||:::|||:::   :::|
    345 s............CysAsnHisHisGlyValCysAsnAsnLysLysAsnC 358

   2270 GCATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCA 2319
        ||   ||||||   |||:::         ||||||:::
    358 ysHisCysAspProThrTyrLeuProProAspCysLys............ 370

   2320 GTTAGGAACCTTCACCCCCCCAAGGATGAAGGACCCAAGGGT 2361
                       |||||||:::   |||   |||
    371 ..............ArgMetLysAspSerTyrProGlyGly 379

Name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-765-243-4

Seq_documentation_block:
; Sequence 4, Application US/08765243
; Patent No. 5935578
;  GENERAL INFORMATION:
;    APPLICANT:  ALVES, KENNETH
;    APPLICANT:  GUPTA, SUNIL K.
;    APPLICANT:  HOLLIS, GREGORY F.
;    TITLE OF INVENTION:  CONTRACEPTIVE VACCINE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MARY A. APPOLLINA
;      STREET:  P.O. BOX 2000, 126 E. LINCOLN AVENUE
;      CITY:  RAHWAY
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07065
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/765,243
;      FILING DATE:
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  APPOLLINA, MARY A
;      REGISTRATION NUMBER:  34,087
;      REFERENCE/DOCKET NUMBER:  19244Y
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)594-3462
;      TELEFAX:  (908)594-4720
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  457 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-765-243-4


alignment_scores:
            Quality:  517.00              Length:    414
              Ratio:  2.163                Gaps:     16
 Percent Similarity:  57.729   Percent Identity: 31.159

alignment_block:
 US-09-634-252A-2 x US-08-765-243-4   ..
```

```
Align seg 1/1  to: US-08-765-243-4  from: 1  to: 457

   1177 ACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTC 1226
        |||::::::   |||      :::   :::|||   |||||||||:::
      9 ThrTyrGlnGlyLysMetCysAspLysAsnTyrAlaGlyGlyValAlaLe 25

   1227 TCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGG 1276
         :::   :::::::|||   ::::::                   ::::::|
     25 uHisProLysAlaValThrLeuGlu.................SerLeuA 36

   1277 CACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAA 1326
        ||   :::|||   |||  |||:::   ::::::|||:::   ::::::
     36 laIleIleLeuValGlnLeuLeuSerLeuSerMetGlyLeuAlaTyrAsp 52

   1327 CCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGC.. 1374
             :::         ||||||:::|||              |||
     53 AspValAsn.........LysCysGlnCys..............GlyVa 61

   1375 .......TGCATCATG......GAGGAAACAGGGGTGTCCCATTCTCGAA 1411
             |||:::|||      |||           |||      |||
     61 lProValCysValMetAsnProGluAlaProHisSerSerGlyValArgA 78

   1412 AATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGA 1461
          ||||||:::|||||||:::   ::::::      |||:::
     78 laPheSerAsnCysSerMetGluAspPheSerLysPheIleThrSerGln 94

   1462 GGTGGAGCCTGCCTTTTCAACAGGCCAACA......AAGCTATTTGAGCC 1505
        ::::::   ||||||   |||:::||||||             ::::::
     95 SerSerHisCysLeuGlnAsnGlnProThrLeuGlnProSerTyrLysMe 111

   1506 CACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTG 1555
         :::   ||||||||||||   ||||||      |||   ||||||||||
    111 tAlaValCysGlyAsnGlyGluValGluGluAspGluIleCysAspCysG 128

   1556 GTTTTCAT......GTGGAATGCTATGGATTATGCTGTAAG......AAA 1593
        ||            :::|||            ||||||:::
    128 lyLysLysGlyCysAlaGluMetProProProCysCysAsnProAspThr 144

   1594 TGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAA 1643
        |||:::||||||:::|||:::::::||||||:::|||   |||||||||
    145 CysLysLeuSerAspGlySerGluCysSerSerGlyIleCysCysAsn.. 160

   1644 TACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGA 1693
             ||||||   ::::::   :::|||      ||||||   |||   :
    161 ....SerCysLysLeuLysArgLysGlyGluValCysArgLeuAlaGlnA 176

   1694 ACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCA 1743
        ::|||||||||:::|||||||||||||:::|||   |||      |||
    176 spGluCysAspValThrGluTyrCysAsnGlyThrSerGluValCys... 191

   1744 CCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCG 1793
           ::::::      |||:::|||:::   |||:::::::::::::
    192 GluAspPhePheValGlnAsnGlyHisProCysAspAsnArgLysTrpIl 208

   1794 CTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCT 1843
         |||   ||||||   |||::::::   ::::::|||||||||   ::::
    208 eCysIleAsnGlyThrCysGlnSerGlyGluGlnGlnCysGlnAspLeuP 225

   1844 GGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAAT 1893
        ::|||      |||:::   :::::::::   |||:::   :::||||||
    225 heGlyIleAspAlaGlyPheGlySerSerGluCysPheTrpGluLeuAsn 241

   1894 ACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGAT 1943
        :::::::::      :::|||:::||||||   ::::::      :::
    242 SerLysSerAspIleSerGlySerCysGlyIleSerAlaGlyGlyTyrLy 258

   1944 TCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGT..... 1986
         :::|||      ::|||   :::||||||   ::::::|||
    258 sGluCysProProAsnAspArgMetCysGlyLysIleIleCysLysTyrG 275

   1987 ..............ACCAATCTTACTCGAGCTCCACGTATT...GGTCAA 2019
                       :::|||      |||      |||   ::::::
    275 lnSerGluAsnIleLeuLysLeuArgSerAlaThrValIleTyrAlaAsn 291
```

```
    2020 CTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGAT 2069
         ::::::|||::::::      :::      |||   ||||||:::
     292 IleSerGlyHisValCysValSerLeuGluTyrProGlnGlyHis..... 306

    2070 TGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCT 2119
                                      ::::::::::::      :
     307 ..............................AsnGluSerGlnLysMetT 313

    2120 ATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGG 2169
         ::|||:::|||||||||   ||||||   :::   :::|||   ::::::
     313 rpValArgAspGlyThrValCysGlySerAsnLysValCysGlnAsnGln 329

    2170 AAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTC 2219
         ||||||:::            |||:::    . :::|||   |||:::::
     330 LysCysValAlaAspThrPheLeuGlyTyr...AspCysAsnLeuGluLy 345

    2220 CAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCT 2269
         :           |||:::   ||||||||||||:::|||:::   :::|
     345 s...........CysAsnHisHisGlyValCysAsnAsnLysLysAsnC 358

    2270 GCATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCA 2319
         ||   ||||||   |||:::         |||||||:::
     358 ysHisCysAspProThrTyrLeuProProAspCysLys............ 370

    2320 GTTAGGAACCTTCACCCCCCCAAGGATGAAGGACCCAAGGGT 2361
                              |||||||:::   |||   |||
     371 ..............ArgMetLysAspSerTyrProGlyGly 379

seq_name: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US95-07295-4

seq_documentation_block:
; Sequence 4, Application PC/TUS9507295
;  GENERAL INFORMATION:
;    APPLICANT:  ALVES, KENNETH
;    APPLICANT:  GUPTA, SUNIL K.
;    APPLICANT:  HOLLIS, GREGORY F.
;    TITLE OF INVENTION:  CONTRACEPTIVE VACCINE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MARY A. APPOLLINA
;      STREET:  P.O. BOX 2000, 126 E. LINCOLN AVENUE
;      CITY:  RAHWAY .
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07065
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/07295
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  APPOLLINA, MARY A
;      REGISTRATION NUMBER:  34,087
;      REFERENCE/DOCKET NUMBER:  19244Y
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)594-3462
;      TELEFAX:  (908)594-4720
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  457 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
PCT-US95-07295-4


alignment_scores:
```

```
            Quality:  517.00              Length:    414
              Ratio:  2.163                 Gaps:     16
  Percent Similarity:  57.729   Percent Identity: 31.159

alignment_block:
 US-09-634-252A-2 x PCT-US95-07295-4   ..

 Align seg 1/1  to: PCT-US95-07295-4  from: 1  to: 457

    1177 ACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTC 1226
         |||::::::   |||      :::   :::|||   |||||||||:::
       9 ThrTyrGlnGlyLysMetCysAspLysAsnTyrAlaGlyGlyValAlaLe 25

    1227 TCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGG 1276
          :::   :::::::|||   ::::::                    ::::::|
      25 uHisProLysAlaValThrLeuGlu.................SerLeuA 36

    1277 CACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAA 1326
         ||   :::|||   |||   |||:::   ::::::|||:::   ::::::
      36 laIleIleLeuValGlnLeuLeuSerLeuSerMetGlyLeuAlaTyrAsp 52

    1327 CCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGC.. 1374
                ::: .        ||||||:::|||                |||
      53 AspValAsn.........LysCysGlnCys..............GlyVa 61

    1375 .......TGCATCATG......GAGGAAACAGGGGTGTCCCATTCTCGAA 1411
                |||:::|||      |||            |||      |||
      61 lProValCysValMetAsnProGluAlaProHisSerSerGlyValArgA 78

    1412 AATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGA 1461
           ||||||:::||||||:::   ::::::       |||:::
      78 laPheSerAsnCysSerMetGluAspPheSerLysPheIleThrSerGln 94

    1462 GGTGGAGCCTGCCTTTTCAACAGGCCAACA......AAGCTATTTGAGCC 1505
         ::::::   ||||||   |||:::||||||            ::::::
      95 SerSerHisCysLeuGlnAsnGlnProThrLeuGlnProSerTyrLysMe 111

    1506 CACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTG 1555
          :::   ||||||||||||   ||||||      |||   |||||||||
     111 tAlaValCysGlyAsnGlyGluValGluGluAspGluIleCysAspCysG 128

    1556 GTTTTCAT......GTGGAATGCTATGGATTATGCTGTAAG......AAA 1593
         ||           :::|||            ||||||:::
     128 lyLysLysGlyCysAlaGluMetProProProCysCysAsnProAspThr 144

    1594 TGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAA 1643
         |||:::|||||||:::|||:::::||||||:::|||   |||||||||
     145 CysLysLeuSerAspGlySerGluCysSerSerGlyIleCysCysAsn.. 160

    1644 TACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGA 1693
             ||||||   ::::::   :::|||      ||||||   |||   :
     161 ....SerCysLysLeuLysArgLysGlyGluValCysArgLeuAlaGlnA 176

    1694 ACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCA 1743
         ::|||||||||:::||||||||||||:::|||   |||      |||
     176 spGluCysAspValThrGluTyrCysAsnGlyThrSerGluValCys... 191

    1744 CCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCG 1793
            ::::::      |||:::|||:::   |||:::::::::::::
     192 GluAspPhePheValGlnAsnGlyHisProCysAspAsnArgLysTrpIl 208

    1794 CTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCT 1843
          |||   ||||||   |||::::::   ::::::|||||||||   ::::
     208 eCysIleAsnGlyThrCysGlnSerGlyGluGlnGlnCysGlnAspLeuP 225

    1844 GGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAAT 1893
         ::|||      |||:::   :::::::::   |||:::   :::||||||
     225 heGlyIleAspAlaGlyPheGlySerSerGluCysPheTrpGluLeuAsn 241

    1894 ACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGAT 1943
         :::::::::      :::|||:::||||||   ::::::      :::
     242 SerLysSerAspIleSerGlySerCysGlyIleSerAlaGlyGlyTyrLy 258
```

```
    1944 TCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGT..... 1986
          :::|||      :::|||   :::||||||   :::::|||
     258 sGluCysProProAsnAspArgMetCysGlyLysIleIleCysLysTyrG 275

    1987 .............ACCAATCTTACTCGAGCTCCACGTATT...GGTCAA 2019
                       :::|||      |||      |||   ::::::
     275 lnSerGluAsnIleLeuLysLeuArgSerAlaThrValIleTyrAlaAsn 291

    2020 CTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGAT 2069
         ::::::|||::::::     :::      |||   |||||||::
     292 IleSerGlyHisValCysValSerLeuGluTyrProGlnGlyHis..... 306

    2070 TGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCT 2119
                                        :::::::::::      :
     307 ............................AsnGluSerGlnLysMetT 313

    2120 ATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGG 2169
         ::|||:::|||||||||   ||||||   :::   :::|||   ::::::
     313 rpValArgAspGlyThrValCysGlySerAsnLysValCysGlnAsnGln 329

    2170 AAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTC 2219
         ||||||:::          |||:::      :::|||   |||:::::
     330 LysCysValAlaAspThrPheLeuGlyTyr...AspCysAsnLeuGluLy 345

    2220 CAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCT 2269
         :          |||:::   ||||||||||||:::|||:::   :::|
     345 s............CysAsnHisHisGlyValCysAsnAsnLysLysAsnC 358

    2270 GCATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCA 2319
         ||   ||||||   |||:::         ||||||:::
     358 ysHisCysAspProThrTyrLeuProProAspCysLys............ 370

    2320 GTTAGGAACCTTCACCCCCCCAAGGATGAAGGACCCAAGGGT 2361
                            ||||||:::   |||   |||
     371 ..............ArgMetLysAspSerTyrProGlyGly 379
```

seq_name: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:US-08-514-014-2

seq_documentation_block:

```
; Sequence 2, Application US/08514014
; Patent No. 5707829
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John
;    APPLICANT:  Kelleher, Kerry
;    APPLICANT:  Carlin, McKeough
;    TITLE OF INVENTION:  DNA SEQUENCES AND SECRETED PROTEINS
;    TITLE OF INVENTION:  ENCODED THEREBY
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
       ADDRESSEE:  Genetics Institute, Inc. -- Legal Affairs
       STREET:  87 CambridgePark Drive
       CITY:  Cambridge
       STATE:  Massachusetts
       COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
       MEDIUM TYPE:  Floppy disk
       COMPUTER:  IBM PC compatible
       OPERATING SYSTEM:  PC-DOS/MS-DOS
       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
       APPLICATION NUMBER:  US/08/514,014
       FILING DATE:
       CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
       NAME:  Brown, Scott A.
       REGISTRATION NUMBER:  32,724
       REFERENCE/DOCKET NUMBER:  GI6000
;    TELECOMMUNICATION INFORMATION:
       TELEPHONE:  (617) 498-8224
       TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  470 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-514-014-2
```

alignment_scores:

```
           Quality:  397.50              Length:    464
             Ratio:  1.430                 Gaps:     19
 Percent Similarity:  59.914   Percent Identity: 27.802
```

alignment_block:

US-09-634-252A-2 x US-08-514-014-2 ..

Align seg 1/1 to: US-08-514-014-2 from: 1 to: 470

```
     328 GAAATCACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGA 377
         ||||||:::   |||:::::::|||
      43 GluIleValCysProLysLysLeu..................... 50

     378 AAGCCCTTATCACGTTCTT.........GACACAAAGGCAAGACACCAGC 418
                  |||:::|||         :::   |||   ::::::   :
      51 .........HisIleLeuHisLysArgGluIleLysAsnAsnGlnThrG 64

     419 AAAAACATAATAAGGCTGTC...CATCTGGCCCAGGCAAGCTTCCAGATT 465
         ::||||||:::|||          :::      :::::::::::|||:::
      64 luLysHisGlyLysGluGluArgTyrGluProGluValGlnTyrGlnMet 80

     466 GAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTT 515
                   |||:::::::::||||||:::|||      ::::::   ||
      81 IleLeuAsnGlyGluGluIleIleLeuSerLeuGlnLysThrLysHisLe 97

     516 GTTGTCTTCTGATTATGTGGAGATTCACTACGAA.............. 549
         ||||:::   ||||||:::|||      |||:::
      97 uLeuGlyProAspTyrThrGluThrLeuTyrSerProArgGlyGluGluI 114

     550 ..AATGGGAAACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCAT 597
           :::   ||||||:::         :::   ||||||||||||||||
     114 leThrThrLysProGlu.........AsnMetGluHisCysTyrTyrLys 127

     598 GGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAA 647
         |||:::|||  :::   |||:::|||   ::::::::::||||||||::
     128 GlyAsnIleLeuAsnGluLysAsnSerValAlaSerIleSerThrCysAs 144

     648 TGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGC 697
         :||||||:::|||   |||                   |||:::|||:::|
     144 pGlyLeuArgGlyTyrPheThrHisHisHisGlnArgTyrGlnIleLysP 161

     698 CACTAGAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCACATATAATC 747
         |||||:::   :::   :::  ::::::|||         |||   |||
     161 roLeuLysSerThrAspGlu..LysGluHisAlaValPheThrSerAsnG 177

     748 CAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTAT 797
           |||:::         :::         ::::::   :::|||
     177 lnGluGluGlnAspProAlaAsnHisThrCysGlyValLysSerThrAsp 193

     798 GGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAA 847
         ||||||:::   |||:::   :::|||      :::          |||||
     194 GlyLysGlnGlyProIleArgIleSerArgSerLeuLysSerProGluLy 210

     848 GAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAA 897
         |::::::                          ::::::  :::   :::|||
     210 sGluAsp.......................PheLeu.ArgAlaGlnLys 218

     898 TATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCG 947
         |||::::::|||   :::|||   |||:::      ||||||:::::::
     219 TyrIleAspLeuTyrLeuValLeuAspAsnAlaPheTyrLysAsnTyrAs 235

     948 CTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTG 997
         :::::::   :::     ::::::|||:::   :::|||:::||||||:
     235 nGluAsnLeuThrLeuIleArgSerPheValPheAspValMetAsnLeuL 252
```

```
     998 TGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCT 1047
         :::::   ||||||   :::   ::::::::::::|||:::||||||:::
     252 euAsnValIleTyr...AsnThrIleAspValGlnValAlaLeuValGly 267

    1048 GTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGT 1097
         :::|||   |||::::::   |||:::|||   ::::::   :::
     268 MetGluIleTrpSerAspGlyAspLysIleLysValValProSerAlaSe 284

    1098 GCAGATGCTCCATGAGTTCTCAAAATACCGG.............CAGC 1132
                :::   :::|||   :::::::::               ::::
     284 rThrThrPheAspAsnPheLeuArgTrpHisSerSerAsnLeuGlyLysL 301

    1133 GCATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTT 1182
         ::|||   :::||||||         :::|||:::|||   :::::::|||
     301 ysIleHisAspHisAla.........GlnLeuLeuSerGlyIleSerPhe 314

    1183 CACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCAC 1232
         :::   :::|||   :::|||:::      :::::::::||||||   ::
     315 AsnAsnArgArgValGlyLeuAlaAlaSerAsnSerLeuCysSerProSe 331

    1233 AAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCA......ATGGCAGTGG 1276
         :   :::|||:::|||   |||                 :::|||::::
     331 r...SerValAlaValIleGluAlaLysLysLysAsnAsnValAlaLeuV 347

    1277 CACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAA 1326
         ::   |||:::|||:::::::|||::::::   ||||||:::
     347 alGlyValMetSerHisGluLeuGlyHisValLeuGlyMetProAspVal 363

    1327 CCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGGTG 1376
         |||   :::   |||           |||   :::      |||:::||
     364 ProPheAsnThrLys...........CysProSer......GlySerCy 374

    1377 CATCATGGAGGAA...ACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGT 1423
         |:::|||::::::      :::   :::      :::   ||||||
     374 sValMetAsnGlnTyrLeuSerSerLysProLysAspPheSerThrS 391

    1424 GCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGC 1473
                 :::::::::   :::|||                   |||
     391 erCysArgAlaHisPheGluArgTyrLeuLeuSerGlnLysProLysCys 407

    1474 CTTTTCAACAGG......CCAACAAAGCTATTTGAGCCCACGGAATGTGG 1517
         |||::::::       ||||||::::::::::              |||||
     408 LeuLeuGlnAlaProIleProThrAsnIleMetThrThrProValCysGl 424

    1518 AAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGG 1567
         ||||      :::|||:::|||||||:::|||||||||||          |
     424 yAsnHisLeuLeuGluValGlyGluAspCysAspCysGlySerProLysG 441

    1568 AATGCTATGGATTATGCTGTAAG......AAATGTTCCCTCTCCAACGGG 1611
         |||||   :::||||||||||:::         |||:::|||:::   |||
     441 luCysThrAsnLeuCysCysGluAlaLeuThrCysLysLeuLysProGly 457

    1612 GCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCA 1650
         :::   |||:::   |||   :::   |||:::|||:::
     458 ThrAspCysGly...GlyAspAlaProAsnHisThrThr 469

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-833-823-2

seq_documentation_block:
; Sequence 2, Application US/08833823
; Patent No. 5969093
;  GENERAL INFORMATION:
;    APPLICANT:  Jacobs, Kenneth
;    APPLICANT:  McCoy, John
;    APPLICANT:  Kelleher, Kerry
;    APPLICANT:  Carlin, McKeough
;    TITLE OF INVENTION:  DNA SEQUENCES AND SECRETED PROTEINS
;    TITLE OF INVENTION:  ENCODED THEREBY
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genetics Institute, Inc. -- Legal Affairs
;      STREET:  87 CambridgePark Drive
;      CITY:  Cambridge
;      STATE:  Massachusetts
;      COUNTRY:  USA
;      ZIP:  02140
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/833,823
;      FILING DATE:  10-APR-1997
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/514,014
;      FILING DATE:  11-AUG-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brown, Scott A.
;      REGISTRATION NUMBER:  32,724
;      REFERENCE/DOCKET NUMBER:  GI6000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 498-8224
;      TELEFAX:  (617) 876-5851
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  470 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-833-823-2


alignment_scores:
            Quality:  397.50              Length:    464
              Ratio:  1.430                 Gaps:     19
 Percent Similarity:  59.914   Percent Identity: 27.802

alignment_block:
 US-09-634-252A-2 x US-08-833-823-2   ..

 Align seg 1/1  to: US-08-833-823-2  from: 1  to: 470

     328 GAAATCACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGA 377
         ||||||:::   |||::::::|||
      43 GluIleValCysProLysLysLeu..........................  50

     378 AAGCCCTTATCACGTTCTT.........GACACAAAGGCAAGACACCAGC 418
                    |||:::|||         :::   |||   ::::::   :
      51 .........HisIleLeuHisLysArgGluIleLysAsnAsnGlnThrG  64

     419 AAAAACATAATAAGGCTGTC...CATCTGGCCCAGGCAAGCTTCCAGATT 465
         ::||||||:::|||       :::      :::::::::::|||:::
      64 luLysHisGlyLysGluGluArgTyrGluProGluValGlnTyrGlnMet 80

     466 GAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTT 515
                  |||:::::::::||||||:::|||      ::::::   ||
      81 IleLeuAsnGlyGluGluIleIleLeuSerLeuGlnLysThrLysHisLe 97

     516 GTTGTCTTCTGATTATGTGGAGATTCACTACGAA.............. 549
         ||||:::   ||||||:::|||      |||:::
      97 uLeuGlyProAspTyrThrGluThrLeuTyrSerProArgGlyGluGluI 114

     550 ..AATGGGAAACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCAT 597
           :::   ||||||:::         :::   ||||||||||||||||
     114 leThrThrLysProGlu.........AsnMetGluHisCysTyrTyrLys 127

     598 GGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAA 647
         |||:::|||   :::   |||::|||   :::::::::||||||||::
     128 GlyAsnIleLeuAsnGluLysAsnSerValAlaSerIleSerThrCysAs 144

     648 TGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGC 697
         :||||||:::|||   |||                   |||:::|||:::|
     144 pGlyLeuArgGlyTyrPheThrHisHisHisGlnArgTyrGlnIleLysP 161
```

```
 698 CACTAGAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCACATATAATC 747
     |||||:::   :::   :::  :::::::|||         |||   |||
 161 roLeuLysSerThrAspGlu..LysGluHisAlaValPheThrSerAsnG 177

 748 CAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTAT 797
       |||:::         :::         ::::::   :::|||
 177 lnGluGluGlnAspProAlaAsnHisThrCysGlyValLysSerThrAsp 193

 798 GGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAA 847
     |||||||:::   |||:::   :::|||      :::         |||||
 194 GlyLysGlnGlyProIleArgIleSerArgSerLeuLysSerProGluLy 210

 848 GAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAA 897
     |::::::                         :::::: :::   :::|||
 210 sGluAsp.....................PheLeu.ArgAlaGlnLys 218

 898 TATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCG 947
     |||::::::|||   :::|||   |||:::      ||||||::::::::
 219 TyrIleAspLeuTyrLeuValLeuAspAsnAlaPheTyrLysAsnTyrAs 235

 948 CTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTG 997
     :::::::   :::      ::::::|||:::   :::|||:::|||||:
 235 nGluAsnLeuThrLeuIleArgSerPheValPheAspValMetAsnLeuL 252

 998 TGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCT 1047
     :::::   ||||||   :::   ::::::::::::|||:::||||||:::
 252 euAsnValIleTyr...AsnThrIleAspValGlnValAlaLeuValGly 267

1048 GTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGT 1097
     :::|||   |||::::::   |||:::|||   ::::::   :::
 268 MetGluIleTrpSerAspGlyAspLysIleLysValValProSerAlaSe 284

1098 GCAGATGCTCCATGAGTTCTCAAAATACCGG.............CAGC 1132
            :::   :::|||   :::::::::              ::::
 284 rThrThrPheAspAsnPheLeuArgTrpHisSerSerAsnLeuGlyLysL 301

1133 GCATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTT 1182
     ::|||   :::||||||         :::|||:::|||   :::::::|||
 301 ysIleHisAspHisAla.........GlnLeuLeuSerGlyIleSerPhe 314

1183 CACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCAC 1232
     :::   :::|||   :::|||:::      :::::::::||||||   ::
 315 AsnAsnArgArgValGlyLeuAlaAlaSerAsnSerLeuCysSerProSe 331

1233 AAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCA......ATGGCAGTGG 1276
     :   :::|||:::|||   |||                 :::|||::::
 331 r...SerValAlaValIleGluAlaLysLysLysAsnAsnValAlaLeuV 347

1277 CACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAA 1326
     ::   |||:::|||::::::|||:::::   ||||||:::
 347 alGlyValMetSerHisGluLeuGlyHisValLeuGlyMetProAspVal 363

 327 CCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTG 1376
     |||   :::   |||            |||   :::      |||:::||
 364 ProPheAsnThrLys...........CysProSer......GlySerCy 374

1377 CATCATGGAGGAA...ACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGT 1423
     |:::|||::::::      :::   :::      :::   ||||||
 374 sValMetAsnGlnTyrLeuSerSerLysProProLysAspPheSerThrS 391

1424 GCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGC 1473
               :::::::::   :::|||                  |||
 391 erCysArgAlaHisPheGluArgTyrLeuLeuSerGlnLysProLysCys 407

1474 CTTTTCAACAGG......CCAACAAAGCTATTTGAGCCCACGGAATGTGG 1517
     |||::::::         ||||||::::::::::          |||||
 408 LeuLeuGlnAlaProIleProThrAsnIleMetThrThrProValCysGl 424

1518 AAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGG 1567
     ||||      :::|||:::||||||:::||||||||||||         |
 424 yAsnHisLeuLeuGluValGlyGluAspCysAspCysGlySerProLysG 441

1568 AATGCTATGGATTATGCTGTAAG......AAATGTTCCCTCTCCAACGGG 1611
     |||||   :::||||||||||:::         |||:::|||:::   |||
 441 luCysThrAsnLeuCysCysGluAlaLeuThrCysLysLeuLysProGly 457

1612 GCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCA 1650
     :::   |||:::   |||   :::   |||:::|||:::
 458 ThrAspCysGly...GlyAspAlaProAsnHisThrThr 469
```

seq_name: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-08-813-150-2

seq_documentation_block:
; Sequence 2, Application US/08813150
; Patent No. 6069229
; GENERAL INFORMATION:
; APPLICANT: Mueller, Christopher
; APPLICANT: Lebecque, Serge J.E.
; APPLICANT: Liu, Yong-Jun
; APPLICANT: Dowling, Lynette M.
; APPLICANT: Huffine, Constance M.
; APPLICANT: Gorman, Daniel M.
; TITLE OF INVENTION: MAMMALIAN PROTEINASES; OXIDOREDUCTASES;
; TITLE OF INVENTION: RELATED REAGENTS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/813,150
; FILING DATE: 07-MAR-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: SF0693
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-852-9196
; TELEFAX: 650-496-1200
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 470 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-813-150-2

alignment_scores:
Quality: 397.50 Length: 464
Ratio: 1.430 Gaps: 19
Percent Similarity: 59.914 Percent Identity: 27.802

alignment_block:
US-09-634-252A-2 x US-08-813-150-2 ..

Align seg 1/1 to: US-08-813-150-2 from: 1 to: 470

```
 328 GAAATCACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGA 377
     |||||||::   |||::::::|||
  43 GluIleValCysProLysLysLeu..................... 50

 378 AAGCCCTTATCACGTTCTT.........GACACAAAGGCAAGACACCAGC 418
                 |||:::|||         :::   |||   ::::::   :
  51 .........HisIleLeuHisLysArgGluIleLysAsnAsnGlnThrG 64
```

```
 419 AAAAACATAATAAGGCTGTC...CATCTGGCCCAGGCAAGCTTCCAGATT 465
     ::|||||||:::|||         :::      ::::::::::::|||:::
  64 luLysHisGlyLysGluGluArgTyrGluProGluValGlnTyrGlnMet  80

 466 GAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTT 515
           |||:::::::::|||||:::|||      ::::::   ||
  81 IleLeuAsnGlyGluGluIleIleLeuSerLeuGlnLysThrLysHisLe  97

 516 GTTGTCTTCTGATTATGTGGAGATTCACTACGAA..............  549
     ||||:::   ||||||:::|||      |||:::
  97 uLeuGlyProAspTyrThrGluThrLeuTyrSerProArgGlyGluGluI 114

 550 ..AATGGGAAACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCAT 597
       :::   ||||||:::         :::   ||||||||||||||||
 114 leThrThrLysProGlu.........AsnMetGluHisCysTyrTyrLys 127

 598 GGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAA 647
     |||:::|||   :::   |||:::|||   :::::::::||||||||::
 128 GlyAsnIleLeuAsnGluLysAsnSerValAlaSerIleSerThrCysAs 144

 648 TGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGC 697
     :||||||:::|||   |||                |||:::|||:::|
 144 pGlyLeuArgGlyTyrPheThrHisHisHisGlnArgTyrGlnIleLysP 161

 698 CACTAGAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCACATATAATC 747
     |||||:::   :::   :::   :::::|||          |||   |||
 161 roLeuLysSerThrAspGlu..LysGluHisAlaValPheThrSerAsnG 177

 748 CAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTAT 797
       |||:::         :::         ::::::   :::|||
 177 lnGluGluGlnAspProAlaAsnHisThrCysGlyValLysSerThrAsp 193

 798 GGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAA 847
     ||||||:::   |||:::   :::|||      :::          |||||
 194 GlyLysGlnGlyProIleArgIleSerArgSerLeuLysSerProGluLy 210

 848 GAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAA 897
     |::::::                        :::::: :::   :::|||
 210 sGluAsp.....................PheLeu.ArgAlaGlnLys 218

 898 TATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCG 947
     |||::::::|||   :::|||   |||:::      ||||||:::::::
 219 TyrIleAspLeuTyrLeuValLeuAspAsnAlaPheTyrLysAsnTyrAs 235

 948 CTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTG 997
     :::::::   :::      ::::::|||:::   :::|||:::||||||:
 235 nGluAsnLeuThrLeuIleArgSerPheValPheAspValMetAsnLeuL 252

 998 TGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCT 1047
     :::::   ||||||   :::   ::::::::::::|||:::||||||:::
 252 euAsnValIleTyr...AsnThrIleAspValGlnValAlaLeuValGly 267

1048 GTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGT 1097
     :::|||   |||::::::   |||:::|||   ::::::   :::
 268 MetGluIleTrpSerAspGlyAspLysIleLysValValProSerAlaSe 284

1098 GCAGATGCTCCATGAGTTCTCAAAATACCGG..............CAGC 1132
           :::   :::|||   ::::::::::               ::::
 284 rThrThrPheAspAsnPheLeuArgTrpHisSerSerAsnProGlyLysL 301

1133 GCATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTT 1182
     ::|||   :::||||||         :::|||:::|||   ::::::|||
 301 ysIleHisAspHisAla.........GlnLeuLeuSerGlyIleSerPhe 314

1183 CACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCAC 1232
     :::   :::|||   :::|||:::      :::::::::||||||  ::
 315 AsnAsnArgArgValGlyLeuAlaAlaSerAsnSerLeuCysSerProSe 331

1233 AAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCA......ATGGCAGTGG 1276
     :   :::|||:::|||   |||                 :::|||::::
 331 r...SerValAlaValIleGluAlaLysLysLysAsnAsnValAlaLeuV 347

1277 CACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAA 1326
```

```
       ::   |||:::|||::::::|||:::::::   ||||||:::
  347 alGlyValMetSerHisGluLeuGlyHisValLeuGlyMetProAspVal 363

 1327 CCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTG 1376
      |||   :::   |||             |||   :::      |||:::||
  364 ProPheAsnThrLys............CysProSer......GlySerCy 374

 1377 CATCATGGAGGAA...ACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGT 1423
      |:::|||::::::      :::   :::      :::   ||||||
  374 sValMetAsnGlnTyrLeuSerSerLysPheProLysAspPheSerThrS 391

 1424 GCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGC 1473
                 :::::::::   :::|||                  |||
  391 erCysArgAlaHisPheGluArgTyrLeuLeuSerGlnLysProLysCys 407

 1474 CTTTTCAACAGG......CCAACAAAGCTATTTGAGCCCACGGAATGTGG 1517
      |||::::::         ||||||:::::::::          |||||
  408 LeuLeuGlnAlaProIleProThrAsnIleMetThrThrProValCysGl 424

 1518 AAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGG 1567
      ||||       :::|||:::||||||:::||||||||||||          |
  424 yAsnHisLeuLeuGluValGlyGluAspCysAspCysGlySerProLysG 441

 1568 AATGCTATGGATTATGCTGTAAG......AAATGTTCCCTCTCCAACGGG 1611
      |||||   :::||||||||:::         |||:::|||:::   |||
  441 luCysThrSerLeuCysCysGluAlaLeuThrCysLysLeuLysProGly 457

 1612 GCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCA 1650
      :::   |||:::   |||   :::   |||:::|||:::
  458 ThrAspCysGly...GlyAspAlaProAsnHisThrThr 469
```

seq_name: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:US-08-655-345-4

```
seq_documentation_block:
; Sequence 4, Application US/08655345
; Patent No. 5830742
;  GENERAL INFORMATION:
;    APPLICANT:  Black, Roy A.
;    APPLICANT:  Rauch, Charles
;    APPLICANT:  March, Carl J.
;    APPLICANT:  Cerretti, Douglas P.
;    TITLE OF INVENTION:  TNF-a CONVERTING ENZYME
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Immunex Corporation
;      STREET:  51 University Street
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Apple Operating System 7.5.2
;      SOFTWARE:  Microsoft Word for Apple, Version 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/655,345
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/504,614
;      FILING DATE:  20-JUL-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/428,458
;      FILING DATE:  8-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Malaska, Stephen L.
;      REGISTRATION NUMBER:  32,655
;      REFERENCE/DOCKET NUMBER:  2507-B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 587-0430
;      TELEFAX:  (206) 233-0644
;  INFORMATION FOR SEQ ID NO:  4:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  681 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-655-345-4


alignment_scores:
            Quality:  353.50              Length:    755
              Ratio:  1.019                Gaps:     42
 Percent Similarity:  45.960   Percent Identity: 22.119

alignment_block:
 US-09-634-252A-2 x US-08-655-345-4   ..

 Align seg 1/1  to: US-08-655-345-4  from: 1  to: 681

     217 CCGCATTGGAATGAAACTGCAGAAAAAAATTTGGGAGTCCTGGCAGATGA 266
         ||||||      :::      ||||||      ::::::|||:::|||
      11 ProHis......GlnArgLeuGluLysLeuAspSerLeuLeuSerAspTy 25

     267 AGACAATACATTGCAACAGAATAGCAGCAGTAATATCAGTTACAGCAATG 316
          |||                                :::|||   ||||||
      25 rAsp.........................IleLeuSerLeuSerAsnI 33

     317 CAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACATCAAC 366
         :::|||   ::::::                                :::
      33 leGlnGlnHisSerValArg.........................Lys 40

     367 CAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCA 416
         :::|||   ::::::      ||||||           :::
      41 ArgAspLeuGlnThrSerThrHisValGluThrLeuLeuThrPheSerAl 57

     417 GCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTG 466
             |||::::::      ::::::|||:::::::::::::        |
      57 aLeuLysArgHisPheLysLeuTyrLeuThrSerSerThr.........G 71

     467 AAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTG 516
         ||   |||:::::::::|||         :::::::::   :::|||
      71 luArgPheSerGlnAsnPhe......LysValValValValAspGlyLys 85

     517 TTGTCTTCTGATTATGTGGAGATTCACTACGAAAATGGGAAACCACAGTA 566
            :::|||:::|||:::   :::   :::::::::
      86 AsnGluSerGluTyrThr...ValLysTrpGlnAsp.............. 96

     567 CTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCA 616
                             ::::::   |||   :::   |||
      97 ......................PhePheThrGlyHisValValGlyGluP 106

     617 AAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTT 666
           ||||||:::|||                          |||      :::
     106 roAspSerArgVal.....................LeuAlaHisIleArg 115

     667 GAAGATGATACCTTCGTG.......................TATATGAT 692
         :::||||||:::::::::                        |||   ||
     116 AspAspAspValIleIleArgIleAsnThrAspGlyAlaGluTyrAsnIl 132

     693 AGAGCCACTA...GAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCAC 739
         ||||||||||   ::::::|||:::|||   |||
     132 eGluProLeuTrpArgPheValAsnAspThrLys................ 143

     740 ATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAAT 789
                 :::|||   :::            |||:::::::::||||||
     144 ........AspLysArgMetLeuValTyrLysSerGluAspIleLysAsn 157

     790 CTCACT.......................................... 795
         ::::::
     158 ValSerArgLeuGlnSerProLysValCysGlyTyrLeuLysValAspAs 174

     796 ..........ATGGAAAGAGGT......GACCAGTGGCCCTTTCTCTCTG 829
                    ::   :::|||      |||:::   |||
     174 nGluGluLeuLeuProLysGlyLeuValAspArgGluPro.........P 188

     830 AATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGT 879
             :::   :::   |||   ||||||   :::::::||
     188 roGluGluLeuValHisArgValLysArgArgAlaAspPro......... 201

     880 ATATTTGAAGAAATGAAA...TATTTGGAACTTATGATTGTTAATGATCA 926
             :::   ||||||         :::|||::::::|||   |||||
     202 ......AspProMetLysAsnThrCysLysLeuLeuValValAlaAspHi 216

     927 CAAAACGTATAAG...AAGCATCGCTCTTCTCATGCACATACCAACAACT 973
         |:::   |||:::         |||::::::::::::   |||:::|||:
     216 sArgPheTyrArgTyrMetGlyArgGlyGluGluSerThrThrThrAsnT 233

     974 TTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTC 1023
         ::      ::::::::::::   ||||||:::||||||:::
     233 yrLeuIleGluLeuIleAspArgValAspAspIleTyrArg......... 246

    1024 AACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGG.............. 1059
         ||||||                        :::|||
     247 AsnThr.......................SerTrpAspAsnAlaGlyPh 255

    1060 .............ACTGAGAAGGATCAGATTGACATCACCACCAACCCT. 1095
                            :::   :::||||||   |||      :::|||
     255 eLysGlyTyrGlyIleGlnIleGluGlnIleArgIleLeuLysSerProG 272

    1095 ..................................................... 1095

     272 lnGluValLysProGlyGluLysHisTyrAsnMetAlaLysSerTyrPro 288

    1096 .......................GTGCAGATGCTCCATGAGTTCTCAAA 1121
                                 |||:::||||||   |||      ::
     289 AsnGluGluLysAspAlaTrpAspValLysMetLeuLeuGlu......Gl 303

    1122 ATACCGGCAGCGCATTAAGCAGCATGCTGATGCTGTG.........CACC 1162
         ::::         |||   ::::::|||:::   |||         ||||
     303 nPheSerPheAspIleAlaGluGluAlaSerLysValCysLeuAlaHisL 320

    1163 TCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTAC... 1209
         ||:::::::         |||                :::|||:::|||
     320 euPheThrTyrGlnAspPheAspMetGlyThrLeuGlyLeuAlaTyrVal 336

    1210 ....................TTTGGAGGTGTCTGTTCTCGCACA..... 1233
                               ||||||||||||   ::::::
     337 GlySerProArgAlaAsnSerHisGlyGlyValCysProLysAlaTyrTy 353

    1234 ................AGAGGAGTTGGTGTGAAT................ 1251
                           :::::::::   :::|||
     353 rSerProValGlyLysLysAsnIleTyrLeuAsnSerGlyLeuThrSerT 370

    1252 .....GAGTATGGTCTTCCAATGGCAGTGGCACAA.........GTATTA 1287
              :::||||||      :::   :::   :::          |||
     370 hrLysAsnTyrGlyLysThrIleLeuThrLysGluAlaAspLeuValThr 386

    1288 TCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAG 1337
         ::::::::::|||:::::::|||:::|||   :::   :::|||::::::
     387 ThrHisGluLeuGlyHisAsnPheGlyAlaGluHisAspProAspGly.. 402

    1338 AAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGC...TGCATCATGG 1384
                ::::|||      :::|||:::   ||||||      :::|||
     403 .LeuAlaGluCysAlaProAsnGluAspGlnGlyGlyLysTyrValMetT 419

    1385 AGGAAACAGGGGTGTCC............CATTCTCGAAAATTTTCAAAG 1422
                ::::||||||            :::::::::   ||||||:::
     419 yrProIleAlaValSerGlyAspHisGluAsnAsnLysMetPheSerAsn 435

    1423 TGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTG 1472
         ||||||      :::            ::::::      :::      ||
     436 CysSerLysGlnSerIleTyrLysThrIleGluSerLysAlaGlnGluCy 452

    1473 CCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATG 1522
         |:::   :::|||   :::|||:::                |||||||||:
     452 sPheGlnGluArgSerAsnLysVal.............CysGlyAsnS 464
```

```
 1523 GATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAAIGC 1572
      ::   |||:::   |||||||||||||||   |||:::
  464 erArgValAspGluGlyGluGluCysAspProGlyIleMetTyrLeuAsn 480

 1573 TATGGATTATGCTGTAAGAAA...TGTTCCCTCTCCAACGGGGCTCACTG 1619
             ||||||::::::   |||:::|||::::::|||::::::||
  481 AsnAspThrCysCysAsnSerAspCysThrLeuLysGluGlyValGlnCy 497

 1620 CAGCGAC......GGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGC 1663
      |||||||      :::|||||||||:::|||      |||   |||:::
  497 sSerAspArgAsnSerProCysCysLysAsn......CysGlnPheGluT 512

 1664 CACGAGGGTATGAATGCCGGGATGCTGTGAAC...GAGTGTGATATTACT 1710
                :::|||::::::|||:::|||      |||      :::
  512 hrAlaGlnLysLysCysGlnGluAlaIleAsnAlaThrCysLysGlyVal 528

  711 GAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCA 1760
      :::||||||||||||:::|||::::::|||||||||      :::   ::
  529 SerTyrCysThrGlyAsnSerSerGluCysProProProGlyAsnAlaGl 545

 1761 AGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGT 1810
      :|||      :::|||      :::   |||:::|||   :::|||:::|
  545 uAspAspThrValCys...LeuAspLeuGlyLysCysLysAspGlyLysC 561

 1811 GCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCA 1860
      ||
  561 ys..................................................  561

 1861 GGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAA 1910
                |||||   ::::::   :::   |||
  562 ......IleProPheCysGluArgGluGlnGlnLeuGlu...........  572

 1911 GGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATG 1960
          :::|||:::                      |||::::::   |
  573 ....SerCysAla.....................CysAsnGluThrA 580

 1961 ATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGT 2010
      ||      |||      :::   |||   :::|||:::
  580 spAsnSerCysLys...ValCysCysArgAspLeuSer............  591

 2011 ATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGG 2060
                   |||:::   :::|||
  592 ..............GlyArgCysValPro...................  596

 2061 CCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACGG 2110
                                      :::|||   :::
  597 ..............................TyrValAspAlaGluGlnL 603

 2111 ATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGT 2160
             :::::::::   |||   ||||||      ::::::
  603 ysAsnLeuPheLeuArgLysGlyLysProCysThrValGlyPheCysAsp 619

 2161 TTAGATCGGAAGTGC 2175
      ::::::   ||||||
  620 MetAsnGlyLysCys 624

seq_name: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:US-09-183-275-4

seq_documentation_block:
; Sequence 4, Application US/09183275
; Patent No. 6013466
;  GENERAL INFORMATION:
;    APPLICANT:  Black, Roy A.
;    APPLICANT:  Rauch, Charles
;    APPLICANT:  March, Carl J.
;    APPLICANT:  Cerretti, Douglas P.
;    TITLE OF INVENTION:  TNF-a CONVERTING ENZYME
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Immunex Corporation
;      STREET:  51 University Street
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  Apple Macintosh
;      OPERATING SYSTEM:  Apple Operating System 7.5.2
;      SOFTWARE:  Microsoft Word for Apple, Version 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/183,275
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/655,345
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/504,614
;      FILING DATE:  20-JUL-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/428,458
;      FILING DATE:  8-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Malaska, Stephen L.
;      REGISTRATION NUMBER:  32,655
;      REFERENCE/DOCKET NUMBER:  2507-B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 587-0430
;      TELEFAX:  (206) 233-0644
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  681 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-183-275-4


alignment_scores:
           Quality:  353.50            Length:    755
             Ratio:  1.019               Gaps:     42
 Percent Similarity:  45.960   Percent Identity: 22.119

alignment_block:
 US-09-634-252A-2 x US-09-183-275-4   ..

 Align seg 1/1  to: US-09-183-275-4  from: 1  to: 681

     217 CCGCATTGGAATGAAACTGCAGAAAAAAATTTGGGAGTCCTGGCAGATGA 266
         ||||||      :::      ||||||      :::::::|||:::|||
      11 ProHis......GlnArgLeuGluLysLeuAspSerLeuLeuSerAspTy 25

     267 AGACAATACATTGCAACAGAATAGCAGCAGTAATATCAGTTACAGCAATG 316
          |||                                ::::||   ||||||
      25 rAsp.........................IleLeuSerLeuSerAsnI 33

     317 CAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACATCAAC 366
           :::|||   ::::::                                :::
      33 leGlnGlnHisSerValArg.........................Lys 40

     367 CAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCA 416
         :::|||   ::::::      ||||||           :::
      41 ArgAspLeuGlnThrSerThrHisValGluThrLeuLeuThrPheSerAl 57

     417 GCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTG 466
             |||::::::      :::::::|||:::::::::::          |
      57 aLeuLysArgHisPheLysLeuTyrLeuThrSerSerThr.........G 71

     467 AAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTG 516
         ||   |||:::::::::|||         :::::::::   :::|||
      71 luArgPheSerGlnAsnPhe......LysValValValValAspGlyLys 85

     517 TTGTCTTCTGATTATGTGGAGATTCACTACGAAAATGGGAAACCACAGTA 566
            :::|||:::|||:::   :::   :::::::::
      86 AsnGluSerGluTyrThr...ValLysTrpGlnAsp.............. 96
```


567 CTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCA 616
97 PhePheThrGlyHisValValGlyGluP 106
617 AAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTT 666
106 roAspSerArgVal.....LeuAlaHisIleArg 115
667 GAAGATGATACCTTCGTG.....TATATGAT 692
116 AspAspAspValIleIleArgIleAsnThrAspGlyAlaGluTyrAsnIl 132
693 AGAGCCACTA...GAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCAC 739
132 eGluProLeuTrpArgPheValAsnAspThrLys..... 143
740 ATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAAT 789
144 AspLysArgMetLeuValTyrLysSerGluAspIleLysAsn 157
790 CTCACT..... 795
158 ValSerArgLeuGlnSerProLysValCysGlyTyrLeuLysValAspAs 174
796 ATGGAAAGAGGT.....GACCAGTGGCCCTTTCTCTCTG 829
174 nGluGluLeuLeuProLysGlyLeuValAspArgGluPro.....P 188
830 AATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGT 879
188 roGluGluLeuValHisArgValLysArgArgAlaAspPro..... 201
880 ATATTTGAAGAAATGAAA...TATTTGGAACTTATGATTGTTAATGATCA 926
202 AspProMetLysAsnThrCysLysLeuLeuValValAlaAspHi 216
927 CAAAACGTATAAG...AAGCATCGCTCTTCTCATGCACATACCAACAACT 973
216 sArgPheTyrArgTyrMetGlyArgGlyGluGluSerThrThrThrAsnT 233
974 TTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTC 1023
233 yrLeuIleGluLeuIleAspArgValAspAspIleTyrArg..... 246
1024 AACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGG..... 1059
247 AsnThr.....SerTrpAspAsnAlaGlyPh 255
1060 ACTGAGAAGGATCAGATTGACATCACCACCAACCCT. 1095
255 eLysGlyTyrGlyIleGlnIleGluGlnIleArgIleLeuLysSerProG 272
095 1095
272 lnGluValLysProGlyGluLysHisTyrAsnMetAlaLysSerTyrPro 288
1096 GTGCAGATGCTCCATGAGTTCTCAAA 1121
289 AsnGluGluLysAspAlaTrpAspValLysMetLeuLeuGlu.....Gl 303
1122 ATACCGGCAGCGCATTAAGCAGCATGCTGATGCTGTG.....CACC 1162
303 nPheSerPheAspIleAlaGluGluAlaSerLysValCysLeuAlaHisL 320
1163 TCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTAC... 1209
320 euPheThrTyrGlnAspPheAspMetGlyThrLeuGlyLeuAlaTyrVal 336
1210 TTTGGAGGTGTCTGTTCTCGCACA..... 1233
337 GlySerProArgAlaAsnSerHisGlyGlyValCysProLysAlaTyrTy 353
1234 AGAGGAGTTGGTGTGAAT..... 1251
353 rSerProValGlyLysLysAsnIleTyrLeuAsnSerGlyLeuThrSerT 370
1252 GAGTATGGTCTTCCAATGGCAGTGGCACAA.....GTATTA 1287
370 hrLysAsnTyrGlyLysThrIleLeuThrLysGluAlaAspLeuValThr 386
1288 TCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAG 1337
387 ThrHisGluLeuGlyHisAsnPheGlyAlaGluHisAspProAspGly.. 402
1338 AAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGC...TGCATCATGG 1384
403 .LeuAlaGluCysAlaProAsnGluAspGlnGlyGlyLysTyrValMetT 419
1385 AGGAAACAGGGGTGTCC.....CATTCTCGAAAATTTTCAAAG 1422
419 yrProIleAlaValSerGlyAspHisGluAsnAsnLysMetPheSerAsn 435
1423 TGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTG 1472
436 CysSerLysGlnSerIleTyrLysThrIleGluSerLysAlaGlnGluCy 452
1473 CCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATG 1522
452 sPheGlnGluArgSerAsnLysVal.....CysGlyAsnS 464
1523 GATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGC 1572
464 erArgValAspGluGlyGluGluCysAspProGlyIleMetTyrLeuAsn 480
1573 TATGGATTATGCTGTAAGAAA...TGTTCCCTCTCCAACGGGGCTCACTG 1619
481 AsnAspThrCysCysAsnSerAspCysThrLeuLysGluGlyValGlnCy 497
1620 CAGCGAC.....GGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGC 1663
497 sSerAspArgAsnSerProCysCysLysAsn.....CysGlnPheGluT 512
1664 CACGAGGGTATGAATGCCGGGATGCTGTGAAC...GAGTGTGATATTACT 1710
512 hrAlaGlnLysLysCysGlnGluAlaIleAsnAlaThrCysLysGlyVal 528
1711 GAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCA 1760
529 SerTyrCysThrGlyAsnSerSerGluCysProProProGlyAsnAlaGl 545
1761 AGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGT 1810
545 uAspAspThrValCys...LeuAspLeuGlyLysCysLysAspGlyLysC 561
1811 GCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCA 1860
561 ys..... 561
1861 GGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAA 1910
562 IleProPheCysGluArgGluGlnGlnLeuGlu..... 572
1911 GGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATG 1960
573 SerCysAla.....CysAsnGluThrA 580
1961 ATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGT 2010
580 spAsnSerCysLys...ValCysCysArgAspLeuSer..... 591
2011 ATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGG 2060
592 GlyArgCysValPro..... 596
2061 CCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACGG 2110

```
                                        :::|||   :::
    597 ..............................TyrValAspAlaGluGlnL 603

    2111 ATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGT 2160
                :::::::::   |||   ||||||      ::::::
     603 ysAsnLeuPheLeuArgLysGlyLysProCysThrValGlyPheCysAsp 619

    2161 TTAGATCGGAAGTGC 2175
         ::::::   ||||||
     620 MetAsnGlyLysCys 624

seq_name: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:PCT-US96-08407-4

seq_documentation_block:
  Sequence 4, Application PC/TUS9608407
  GENERAL INFORMATION:
    APPLICANT:  Immunex Corporation
    TITLE OF INVENTION:  TNF-a CONVERTING ENZYME
    NUMBER OF SEQUENCES:  9
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Immunex Corporation
      STREET:  51 University Street
      CITY:  Seattle
      STATE:  WA
      COUNTRY:  USA
      ZIP:  98101
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Floppy disk
      COMPUTER:  Apple Macintosh
      OPERATING SYSTEM:  Apple Operating System 7.5.2
      SOFTWARE:  Microsoft Word for Apple, Version 6.0
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  PCT/US96/08407
      FILING DATE:  03-JUN-1996
      CLASSIFICATION:
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  --to be assigned--
      FILING DATE:  23-MAY-1996
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  08/504,614
      FILING DATE:  20-JUL-1995
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  08/428,458
      FILING DATE:  8-JUN-1995
    ATTORNEY/AGENT INFORMATION:
      NAME:  Malaska, Stephen L.
      REGISTRATION NUMBER:  32,655
      REFERENCE/DOCKET NUMBER:  2507-WO
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  (206) 587-0430
      TELEFAX:  (206) 233-0644
  INFORMATION FOR SEQ ID NO:  4:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  681 amino acids
      TYPE:  amino acid
      TOPOLOGY:  linear
    MOLECULE TYPE:  protein
PCT-US96-08407-4


alignment_scores:
            Quality:  353.50              Length:    755
              Ratio:  1.019                 Gaps:     42
 Percent Similarity:  45.960   Percent Identity: 22.119

alignment_block:
 US-09-634-252A-2 x PCT-US96-08407-4    ..

 Align seg 1/1  to: PCT-US96-08407-4  from: 1  to: 681

     217 CCGCATTGGAATGAAACTGCAGAAAAAAATTTGGGAGTCCTGGCAGATGA 266
         ||||||      :::      ||||||      :::::::|||:::|||
      11 ProHis......GlnArgLeuGluLysLeuAspSerLeuLeuSerAspTy 25

     267 AGACAATACATTGCAACAGAATAGCAGCAGTAATATCAGTTACAGCAATG 316
          |||                              :::|||   ||||||
      25 rAsp..........................IleLeuSerLeuSerAsnI 33

     317 CAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACATCAAC 366
         :::|||   ::::::                                :::
      33 leGlnGlnHisSerValArg..........................Lys 40

     367 CAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCA 416
         :::|||   ::::::      ||||||          :::
      41 ArgAspLeuGlnThrSerThrHisValGluThrLeuLeuThrPheSerAl 57

     417 GCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTG 466
          |||::::::      ::::::|||:::::::::::::         |
      57 aLeuLysArgHisPheLysLeuTyrLeuThrSerSerThr.........G 71

     467 AAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTG 516
         ||   |||::::::::::|||         :::::::::   :::|||
      71 luArgPheSerGlnAsnPhe......LysValValValValAspGlyLys 85

     517 TTGTCTTCTGATTATGTGGAGATTCACTACGAAAATGGGAAACCACAGTA 566
         :::|||:::|||:::   :::   :::::::::
      86 AsnGluSerGluTyrThr...ValLysTrpGlnAsp.............. 96

     567 CTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCA 616
                              ::::::   |||   :::   |||
      97 .....................PhePheThrGlyHisValValGlyGluP 106

     617 AAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTT 666
           ||||||:::|||                |||      :::
     106 roAspSerArgVal.....................LeuAlaHisIleArg 115

     667 GAAGATGATACCTTCGTG.......................TATATGAT 692
         :::||||||:::::::::                      |||   ||
     116 AspAspAspValIleIleArgIleAsnThrAspGlyAlaGluTyrAsnIl 132

     693 AGAGCCACTA...GAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCAC 739
         ||||||||||   :::::::|||:::|||   |||
     132 eGluProLeuTrpArgPheValAsnAspThrLys................ 143

     740 ATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAAT 789
               :::|||   :::             |||:::::::::||||||
     144 ........AspLysArgMetLeuValTyrLysSerGluAspIleLysAsn 157

     790 CTCACT.......................................... 795
         ::::::
     158 ValSerArgLeuGlnSerProLysValCysGlyTyrLeuLysValAspAs 174

     796 ..........ATGGAAAGAGGT......GACCAGTGGCCCTTTCTCTCTG 829
                 :::   :::|||      |||:::   |||
     174 nGluGluLeuLeuProLysGlyLeuValAspArgGluPro.........P 188

     830 AATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGT 879
           :::   :::   |||   ||||||   :::::|||
     188 roGluGluLeuValHisArgValLysArgArgAlaAspPro......... 201

     880 ATATTTGAAGAAATGAAA...TATTTGGAACTTATGATTGTTAATGATCA 926
               :::   ||||||        :::|||::::::|||   |||||
     202 ......AspProMetLysAsnThrCysLysLeuLeuValValAlaAspHi 216

     927 CAAAACGTATAAG...AAGCATCGCTCTTCTCATGCACATACCAACAACT 973
         |:::   |||:::         |||:::::::::::::   |||:::|||:
     216 sArgPheTyrArgTyrMetGlyArgGlyGluGluSerThrThrThrAsnT 233

     974 TTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTC 1023
         ::      ::::::::::::   ||||||:::|||||||:::
     233 yrLeuIleGluLeuIleAspArgValAspAspIleTyrArg......... 246

    1024 AACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGG.............. 1059
         ||||||                        :::|||
     247 AsnThr.......................SerTrpAspAsnAlaGlyPh 255
```

```
1060 .............ACTGAGAAGGATCAGATTGACATCACCACCAACCCT. 1095
                      :::   :::||||||   |||      :::|||
 255 eLysGlyTyrGlyIleGlnIleGluGlnIleArgIleLeuLysSerProG 272

1095 ................................................. 1095

 272 lnGluValLysProGlyGluLysHisTyrAsnMetAlaLysSerTyrPro 288

1096 .......................GTGCAGATGCTCCATGAGTTCTCAAA 1121
                            |||:::||||||   |||      ::
 289 AsnGluGluLysAspAlaTrpAspValLysMetLeuLeuGlu......Gl 303

1122 ATACCGGCAGCGCATTAAGCAGCATGCTGATGCTGTG.........CACC 1162
     ::::         |||   ::::::|||:::   |||         ||||
 303 nPheSerPheAspIleAlaGluGluAlaSerLysValCysLeuAlaHisL 320

1163 TCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTAC... 1209
     ||::::::         |||                :::|||:::|||
 320 euPheThrTyrGlnAspPheAspMetGlyThrLeuGlyLeuAlaTyrVal 336

1210 .....................TTTGGAGGTGTCTGTTCTCGCACA..... 1233
                           ||||||||||||   ::::::
 337 GlySerProArgAlaAsnSerHisGlyGlyValCysProLysAlaTyrTy 353

1234 ................AGAGGAGTTGGTGTGAAT................ 1251
                     :::::::::   :::|||
 353 rSerProValGlyLysLysAsnIleTyrLeuAsnSerGlyLeuThrSerT 370

1252 .....GAGTATGGTCTTCCAATGGCAGTGGCACAA.........GTATTA 1287
          :::||||||      :::   :::   :::         |||
 370 hrLysAsnTyrGlyLysThrIleLeuThrLysGluAlaAspLeuValThr 386

1288 TCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAG 1337
     ::::::::::|||:::::::|||:::|||   :::   :::|||::::::
 387 ThrHisGluLeuGlyHisAsnPheGlyAlaGluHisAspProAspGly.. 402

1338 AAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGC...TGCATCATGG 1384
           :::|||      :::|||:::   ||||||      :::|||
 403 .LeuAlaGluCysAlaProAsnGluAspGlnGlyGlyLysTyrValMetT 419

1385 AGGAAACAGGGGTGTCC............CATTCTCGAAAATTTTCAAAG 1422
           :::||||||             :::::::::   ||||||:::
 419 yrProIleAlaValSerGlyAspHisGluAsnAsnLysMetPheSerAsn 435

1423 TGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTG 1472
     ||||||      :::          ::::::      :::       ||
 436 CysSerLysGlnSerIleTyrLysThrIleGluSerLysAlaGlnGluCy 452

1473 CCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATG 1522
     |:::   :::|||   :::|||:::                |||||||||:
 452 sPheGlnGluArgSerAsnLysVal..............CysGlyAsnS 464

 523 GATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGC 1572
     ::   |||:::   ||||||||||||||   |||:::
 464 erArgValAspGluGlyGluGluCysAspProGlyIleMetTyrLeuAsn 480

1573 TATGGATTATGCTGTAAGAAA...TGTTCCCTCTCCAACGGGGCTCACTG 1619
           ||||||::::::   |||:::|||::::::|||::::::||
 481 AsnAspThrCysCysAsnSerAspCysThrLeuLysGluGlyValGlnCy 497

1620 CAGCGAC......GGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGC 1663
     |||||||      :::||||||||:::|||      |||   |||:::
 497 sSerAspArgAsnSerProCysCysLysAsn......CysGlnPheGluT 512

1664 CACGAGGGTATGAATGCCGGGATGCTGTGAAC...GAGTGTGATATTACT 1710
           :::|||::::::|||:::|||      |||      :::
 512 hrAlaGlnLysLysCysGlnGluAlaIleAsnAlaThrCysLysGlyVal 528

1711 GAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCA 1760
     :::||||||||||||:::|||::::::|||||||||      :::   ::
 529 SerTyrCysThrGlyAsnSerSerGluCysProProProGlyAsnAlaGl 545

1761 AGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGT 1810
```

```
     :|||      :::|||      :::   |||:::|||   :::|||:::|
 545 uAspAspThrValCys...LeuAspLeuGlyLysCysLysAspGlyLysC 561

1811 GCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCA 1860
     ||
 561 ys..................................................  561

1861 GGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAA 1910
                |||||   ::::::   :::   |||
 562 ......IleProPheCysGluArgGluGlnGlnLeuGlu..........  572

1911 GGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATG 1960
        :::|||:::                          |||::::::   |
 573 ....SerCysAla.......................CysAsnGluThrA 580

1961 ATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGT 2010
     ||      |||      :::   |||   :::|||:::
 580 spAsnSerCysLys...ValCysCysArgAspLeuSer............ 591

2011 ATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGG 2060
                  |||:::   :::|||
 592 ..............GlyArgCysValPro.....'.............. 596

2061 CCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACGG 2110
                                       :::|||   :::
 597 ............................TyrValAspAlaGluGlnL 603

2111 ATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGT 2160
           :::::::::   |||   ||||||      ::::::
 603 ysAsnLeuPheLeuArgLysGlyLysProCysThrValGlyPheCysAsp 619

2161 TTAGATCGGAAGTGC 2175
     ::::::   ||||||
 620 MetAsnGlyLysCys 624
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2001, 19:30:52 ; Search time 1869.46 Seconds
(without alignments)
12636.099 Million cell updates/sec

Title: US-09-634-252A-2
Perfect score: 2499
Sequence: 1 atgaagccgcccggcagcag.....ctcagcaaggccccatctga 2499

[redacted]ng table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 10228115 seqs, 4726426750 residues

Total number of hits satisfying chosen parameters: 20456230

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 60 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_esthum29:*
177: em_esthum30:*
178: em_esthum31:*
179: em_esthum32:*
180: em_esthum33:*
181: em_estom3:*
182: em_estpl11:*
183: em_estro21:*
184: em_estro22:*
185: em_estro23:*
186: em_htc:*
187: gb_est107:*
188: gb_est108:*
189: gb_est109:*
190: gb_est110:*
191: gb_est111:*
192: gb_htc:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_other:*
207: em_gss_pln1:*
208: em_gss_pln2:*
209: em_gss_pro:*
210: em_gss_rod1:*
211: em_gss_rod2:*
212: em_gss_rod3:*
213: em_gss_rod4:*
214: em_gss_rod5:*
215: em_gss_vrt1:*
216: em_gss_vrt2:*
217: em_gss_vrt3:*
218: gb_gss1:*
219: gb_gss2:*
220: gb_gss3:*
221: gb_gss4:*
222: gb_gss5:*
223: gb_gss6:*
224: gb_gss7:*
225: gb_gss8:*
226: gb_gss9:*
227: gb_gss10:*
228: gb_gss11:*
229: gb_gss12:*
230: gb_gss13:*
231: gb_gss14:*
232: gb_gss15:*
233: gb_gss16:*
234: gb_gss17:*
235: gb_gss18:*
236: gb_gss19:*
237: gb_gss20:*
238: gb_gss21:*
239: gb_gss22:*
240: gb_gss23:*
241: gb_gss24:*
242: gb_gss25:*
243: gb_gss26:*
244: gb_gss27:*
245: gb_gss28:*
246: gb_gss29:*
247: gb_gss30:*
248: gb_gss31:*
249: gb_gss32:*
250: gb_gss33:*
251: gb_gss34:*
252: em_gss_inv4:*
253: em_gss_rod6:*
254: em_gss_rod7:*
255: em_gss_rod8:*
256: gb_gss35:*
257: gb_gss36:*
258: gb_gss37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	639.6	25.6	752	168	BF691008	BF691008 602247043
2	624.8	25.0	698	23	AI692184	AI692184 wd37f10.x
3	551.8	22.1	626	114	AW303311	AW303311 xv17e08.x
4	491.4	19.7	553	122	AW961362	AW961362 EST373434
5	479	19.2	552	119	AW660536	AW660536 99459 MAR
6	413.2	16.5	447	171	BF905823	BF905823 IL3-MT026
7	396.4	15.9	503	147	BF308417	BF308417 601890734
8	336.8	13.5	428	11	AA718688	AA718688 vt87c03.r
9	329.4	13.2	340	157	F08148	F08148 HSC2RF011 n
10	295	11.8	375	8	AA511039	AA511039 vh55g05.r
11	271.2	10.9	405	187	R52569	R52569 yg81c01.r1
12	267	10.7	326	187	R15038	R15038 yf86d08.r1
13	259.8	10.4	763	168	BF693020	BF693020 602080016
14	257	10.3	261	5	AA317222	AA317222 EST19173
c 15	232	9.3	397	122	AW898896	AW898896 CM0-NN007
c 16	228.8	9.2	237	122	AW900312	AW900312 CM4-NN100
c 17	211	8.4	576	19	AI342396	AI342396 qt27g06.x
18	190.6	7.6	504	166	BE309477	BE309477 601095217
19	182.8	7.3	476	22	AI597498	AI597498 vj50a07.x
20	181.4	7.3	468	105	AL135156	AL135156 DKFZp762L
21	175.8	7.0	223	114	AW307806	AW307806 2817 MARC
22	161.8	6.5	456	114	AW326729	AW326729 19656 MAR
23	155.8	6.2	252	138	BE651851	BE651851 UI-M-AL0-
c 24	154.8	6.2	421	21	AI535213	AI535213 UI-R-C3-t
25	141	5.6	675	114	AW280443	AW280443 fj40f01.y
c 26	122.4	4.9	526	1	AA050162	AA050162 mj17b03.r
c 27	120.6	4.8	127	122	AW900291	AW900291 CM4-NN100
c 28	113.8	4.6	531	190	W75581	W75581 me54f09.r1
c 29	108.6	4.3	523	191	W96980	W96980 mf61a03.r1
30	107.2	4.3	214	145	BF190216	BF190216 236705 MA
31	105.6	4.2	562	111	AW125422	AW125422 UI-M-BH2.
32	102.2	4.1	639	114	AW280355	AW280355 fj38g06.y
33	102.2	4.1	783	108	AU136074	AU136074 AU136074
34	100.8	4.0	648	108	AU137049	AU137049 AU137049
35	100.2	4.0	855	108	AU137994	AU137994 AU137994
36	99	4.0	571	172	BF997087	BF997087 QV3-GN020
37	98.6	3.9	660	108	AU139359	AU139359 AU139359
38	98.6	3.9	681	108	AU134844	AU134844 AU134844
39	98.6	3.9	781	108	AU135801	AU135801 AU135801
40	98.6	3.9	802	108	AU140094	AU140094 AU140094
41	98.6	3.9	974	141	BE875422	BE875422 601489141
42	98.4	3.9	254	144	BF090250	BF090250 PM3-NN108
43	98	3.9	730	108	AU138650	AU138650 AU138650
44	97.6	3.9	723	108	AU139616	AU139616 AU139616
45	97.6	3.9	762	108	AU139012	AU139012 AU139012
46	96	3.8	597	108	AU139162	AU139162 AU139162
47	95.8	3.8	120	150	BF543743	BF543743 UI-R-C3-t
48	94.8	3.8	806	108	AU134792	AU134792 AU134792
49	92.8	3.7	621	108	AU134660	AU134660 AU134660
50	92.8	3.7	727	107	AU127125	AU127125 AU127125
c 51	91.6	3.7	500	246	AZ564512	AZ564512 RPCI-23-1
52	91.2	3.6	713	108	AU138921	AU138921 AU138921
c 53	90	3.6	519	172	BF988372	BF988372 MR1-GN017
54	87.2	3.5	604	108	AU137378	AU137378 AU137378
55	84.8	3.4	759	108	AU138290	AU138290 AU138290
56	83	3.3	611	115	AW390456	AW390456 RC3-ST018
57	83	3.3	620	108	AU134899	AU134899 AU134899
58	81.2	3.2	693	108	AU138455	AU138455 AU138455
c 59	80.6	3.2	383	172	BF987229	BF987229 QV0-GN014
c 60	79.8	3.2	239	9	AA566127	AA566127 ESTKB025

ALIGNMENTS

RESULT 1
BF691008

Isolate

```
LOCUS       BF691008      752 bp    mRNA            EST       22-DEC-2000
DEFINITION  602247043F1 NIH_MGC_62 Homo sapiens cDNA clone IMAGE:4332275 5',
            mRNA sequence.
ACCESSION   BF691008
VERSION     BF691008.1  GI:11976416
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 752)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: CLONETECH Laboratories, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1198  row: e  column: 12
            High quality sequence stop: 719.
FEATURES             Location/Qualifiers
     source          1. .752
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4332275"
                     /clone_lib="NIH_MGC_62"
                     /tissue_type="melanotic melanoma, high MDR"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: skin; Vector: pDNR-LIB (Clontech); Site_1:
                     SfiI (ggccgcctcggcc); Site_2: SfiI (ggccattatggcc);
                     Double-stranded cDNA was prepared from cell line RNA.  5'
                     and 3' adaptors were used in cloning as follows: 5'
                     adaptor sequence:  5'-CACGGCCATTATGGCC-3' and 3' adaptor
                     sequence:  5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3'
                     (where B = A, C, or G and N = A, C,  G, or T).   Average
                     insert size 1.75 kb (range 0.9-4.0 kb).  15/15 colonies
                     contained  inserts by PCR. This library was enriched for
                     full-length clones and was constructed by Clontech
                     Laboratories (Palo Alto, CA)."
BASE COUNT      195 a    151 c    217 g    189 t
ORIGIN
```

Query Match 25.6%; Score 639.6; DB 168; Length 752;
Best Local Similarity 97.7%; Pred. No. 2.1e-169;
Matches 713; Conservative 0; Mismatches 9; Indels 8; Gaps 6;

```
Qy   1522 ggatacgtggaagctggggaggagtgtgattgtggttttcatgtggaatgctatggatta 1581
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTATGGATTA 60

Qy   1582 tgctgtaagaaatgttccctctccaacggggctcactgcagcgacgggccctgctgtaac 1641
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGCTGTAAGAAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAAC 120

Qy   1642 aatacctcatgtctttttcagccacgagggtatgaatgccgggatgctgtgaacgagtgt 1701
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGT 180

Qy   1702 gatattactgaatattgtactggagactctggtcagtgcccaccaaatcttcataagcaa 1761
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAA 240

Qy   1762 gacggatatgcatgcaatcaaaatcagggccgctgctacaatggcgagtgcaagaccaga 1821
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGA 300

Qy   1822 gacaaccagtgtcagtacatctggggaacaaaggctgcagggtctgacaagttctgctat 1881
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTAT 360

Qy    1882 gaaaagctgaatacagaaggcactgagaagggaaactgcgggaaggatggagaccggtgg 1941
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGG 420

Qy    1942 attcagtgcagcaaacatgatgtgttctgtggattcttactctgtaccaatcttactcga 2001
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 ATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGA 480

Qy    2002 gct-ccacgtattggtcaacttcagggtgagatcattccaacttccttct-accatcaag 2059
           ||| ||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db     481 GCTCCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTAACCATCAAG 540

Qy    2060 gccgggtgattgactgcagtggtgcccatgtagttttagatgatgatacggatgtgggct 2119
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 GCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCT 600

Qy    2120 atgtagaagatggaacgccatgtggcccgtctatgatgtgtttagatcggaagtgcctac 2179
           |||||||||||||| |||||||||||| |||||||||||||| ||||||||||||||||||||
Db     601 ATGTAGAAGATGGGACGCCATGTGG-CCGTCTATGATGTG-TTAGATCGGAAGTGCCTAC 658

Qy    2180 aaattcaagccctaaatatgagcagctgtccactcgattccaagggtaaagtctgttcgg 2239
           |||||||||||||||||||||||| ||||   |||||||||   |  ||||||||||||
Db     659 AAATTCAAGCCCTAAATATGAGCAG-TGTCACTTCGATTCCA---GGTAAGTCTGTTCGG 714

Qy    2240 gccatggggt 2249
           ||   |||||
Db     715 GCAGGGGGGT 724
```

RESULT 2
AI692184
LOCUS AI692184 698 bp mRNA EST 02-JUN-1999
DEFINITION wd37f10.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2330347 3' similar to TR:O75077 O75077 MDC3. ;contains PTR5.t3 PTR5 repetitive element ;, mRNA sequence.
ACCESSION AI692184
VERSION AI692184.1 GI:4969524
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 698)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT ~~Contact: Robert Strausberg, Ph.D.~~
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 486.

FEATURES Location/Qualifiers
source 1. .698
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2330347"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbHL19W, testis NHT, and B-cell NCI_CGAP_GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 161 a 228 c 178 g 129 t 2 others
ORIGIN

Query Match 25.0%; Score 624.8; DB 23; Length 698;
Best Local Similarity 99.4%; Pred. No. 3.1e-165;
Matches 626; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```
Qy      1 atgaagccgcccggcagcagctcgcggcagccgcccctggcgggctgcagccttgccggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69 ATGAAGCCGCCCGGCAGCAGCTCGCGGCAGCCGCCCCTGGCGGGCTGCAGCCTTGCCGGC 128

Qy     61 gcttcctgcggcccccaacgcggccccgccggctcggtgcctgccagcgccccggcccgc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    129 GCTTCCTGCGGCCCCCAACGCGGCCCCGCCGGCTCGGTGCCTGCCAGCGCCCCGGCCCGC 188

Qy    121 acgccgccctgccgcctgcttctcgtccttctcctgctgcctccgctcgccgcctcgtcc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    189 ACGCCGCCCTGCCGCCTGCTTCTCGTCCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTCC 248

Qy    181 cggccccgcgcctgggggctgctgcgcccagcgctccgcattggaatgaaactgcagaa 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    249 CGGCCCCGCGCCTGGGGGGCTGCTGCGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAA 308

Qy    241 aaaaatttgggagtcctggcagatgaagacaatacattgcaacagaatagcagcagtaat 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    309 AAAAATTTGGGAGTCCTGGCAGATGAAGACAATACATTGCAACAGAATAGCAGCAGTAAT 368

Qy    301 atcagttacagcaatgcaatgcagaaagaaatcacactgccttcaagactcatatattac 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    369 ATCAGTTACAGCAATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTAC 428

Qy    361 atcaaccaagactcggaaagcccttatcacgttcttgacacaaaggcaagacaccagcaa 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    429 ATCAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAA 488

Qy    421 aaacataataaggctgtccatctggcccaggcaagcttccagattgaagccttcggctcc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    489 AAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCC 548

Qy    481 aaattcattcttgacctcatactgaacaatggtttgttgtcttctgattatgtggagatt 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    549 AAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATT 608

Qy    541 cactacgaaaatgggaaaccacagtactctaagggtggagagcactgttactaccatgga 600
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    609 CACTACGNAAATGGGAAACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGA 668

Qy    601 agcatcagaggcgtcaaagactccaaggtg 630
          ||||||||  |||||| ||||||||||||||
Db    669 AGCATCAGTAGCGTCANAGACTCCAAGGTG 698
```

RESULT 3
AW303311
LOCUS AW303311 626 bp mRNA EST 18-JAN-2000
DEFINITION xv17e08.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone IMAGE:2813414 3' similar to TR:O75077 O75077 MDC3. ;, mRNA sequence.
ACCESSION AW303311
VERSION AW303311.1 GI:6712991
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 626)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL ~~Unpublished (1997)~~

```
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: similarity on wrong strand
            Possible reversed clone: polyT not found
            Seq primer: -40UP from Gibco
            High quality sequence stop: 490.
FEATURES             Location/Qualifiers
     source          1. .626
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2813414"
                     /clone_lib="Soares_NFL_T_GBC_S1"
                     /lab_host="DH10B"
                     /note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Equal amounts of plasmid DNA from three normalized
                     libraries (fetal lung NbHL19W, testis NHT, and B-cell
                     NCI_CGAP_GCB1) were mixed, and ss circles were made in
                     vitro. Following HAP purification, this DNA was used as
                     tracer in a subtractive hybridization reaction. The driver
                     was PCR-amplified cDNAs from pools of 5,000 clones made
                     from the same 3 libraries. The pools consisted of
                     I.M.A.G.E. clones 297480-302087, 682632-687239,
                     726408-728711, and 729096-731399. Subtraction by Bento
                     Soares and M. Fatima Bonaldo. "
BASE COUNT      141 a    213 c    156 g    114 t      2 others
ORIGIN
```

```
  Query Match             22.1%;  Score 551.8;  DB 114;  Length 626;
  Best Local Similarity   99.3%;  Pred. No. 1.3e-144;
  Matches  553;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      1 atgaagccgcccggcagcagctcgcggcagccgcccctggcgggctgcagccttgccggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69 ATGAAGCCGCCCGGCAGCAGCTCGCGGCAGCCGCCCCTGGCGGGCTGCAGCCTTGCCGGC 128

Qy     61 gcttcctgcggcccccaacgcggccccgccggctcggtgcctgccagcgccccggcccgc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    129 GCTTCCTGCGGCCCCCAACGCGGCCCCGCCGGCTCGGTGCCTGCCAGCGCCCCGGCCCGC 188

Qy    121 acgccgccctgccgcctgcttctcgtccttctcctgctgcctccgctcgccgcctcgtcc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    189 ACGCCGCCCTGCCGCCTGCTTCTCGTCCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTCC 248

Qy    181 cggccccgcgcctgggggggctgctgcgcccagcgctccgcattggaatgaaactgcagaa 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    249 CGGCCCCGCGCCTGGGGGGCTGCTGCGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAA 308

Qy    241 aaaaatttgggagtcctggcagatgaagacaatacattgcaacagaatagcagcagtaat 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    309 AAAAATTTGGGAGTCCTGGCAGATGAAGACAATACATTGCAACAGAATAGCAGCAGTAAT 368

Qy    301 atcagttacagcaatgcaatgcagaaagaaatcacactgccttcaagactcatatattac 360
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    369 ATCAGTTACAGCAATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTAC 428

Qy    361 atcaaccaagactcggaaagcccttatcacgttcttgacacaaaggcaagacaccagcaa 420
          ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    429 ATCAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAA 488

Qy    421 aaacataataaggctgtccatctggcccaggcaagcttccagattgaagccttcggctcc 480
          |||||||||||||||||||||||||| |||||||||||||||| ||||||| |||||||
Db    489 AAACATAATAAGGCTGTCCATCTGGCCCANGCAAGCTTCCAGATTGAAGCCTTCCGCTCC 548

Qy    481 aaattcattcttgacctcatactgaacaatggtttgttgtcttctgattatgtggagatt 540
          |||||||||||||||||||||||||||||||||| ||||||||| ||||||||||||||||
Db    549 AAATTCATTCTTGACCTCATACTGAACAATGGTTTGNTGTCTTCTGATTATGTGGAGATT 608

Qy    541 cactacgaaaatgggaa 557
          ||||||||||||||| ||
```

```
Db    609 CACTACGAAAATGGAAA 625


RESULT   4
AW961362
LOCUS       AW961362      553 bp    mRNA            EST       01-JUN-2000
DEFINITION  EST373434 MAGE resequences, MAGG Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW961362
VERSION     AW961362.1  GI:8151046
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 553)
  AUTHORS   Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Holt
            ,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
            Quackenbush,J.
  TITLE     Assessment of gene expression patterns in a model of colon tumor
            metastasis using a 19,200 element cDNA microarray
  JOURNAL   Unpublished (2000)
COMMENT     Contact: John Quackenbush
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 3528
            Fax: 301 838 0208
            Email: johnq@tigr.org
            Plate: 157
            Seq primer: Reverse.
FEATURES             Location/Qualifiers
     source          1. .553
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MAGE resequences, MAGG"
                     /note="Vector: pBluescriptSKm"
BASE COUNT      149 a    113 c    148 g    143 t
ORIGIN
```

```
  Query Match             19.7%;  Score 491.4;  DB 122;  Length 553;
  Best Local Similarity   94.2%;  Pred. No. 1.5e-127;
  Matches  521;  Conservative    0;  Mismatches   31;  Indels    1;  Gaps    1;

Qy   1548 tgattgtggttttcatgtggaatgctatggattatgctgtaagaaatgttccctctccaa 1607
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TGATTGTGGTTTTCATGTGGAATGCTATGGATTATGCTGTAAGAAATGTTCCCTCTCCAA 60

Qy   1608 cggggctcactgcagcgacgggccctgctgtaacaatacctcatgtctttttcagccacg 1667
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACG 120

Qy   1668 agggtatgaatgccgggatgctgtgaacgagtgtgatattactgaatattgtactggaga 1727
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 AGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTACTGGAGA 180

Qy   1728 ctctggtcagtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatca 1787
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 CTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCA 240

Qy   1788 gggccgctgctacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctgggg 1847
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GGGCCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGG 300

Qy   1848 aacaaaggctgcagggtctgacaagttctgctatgaaaagctgaatacagaaggcactga 1907
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGA 360

Qy   1908 gaagggaaactgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtgtt 1967
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTT 420

Qy   1968 ctgtggattcttactctgtaccaatcttactcgagctccacgtattggtcaacttcaggg 2027
          |||||||||||||||||||| ||||||||| ||||||      | |  |  || ||||
```

```
Db    421 CTGTGGATTCTTACTCTGTACAAATCTTACTTGAGCTCACGTATTGGTCAACTTTAAGGG 480

Qy   2028 tgagatcattccaacttccttctaccatcaaggccgggtgattgac-tgcagtggtgccc 2086
          |||||    ||| |  | | || |||| | |||||| || ||| ||| ||||||||  |||
Db    481 TGAGACATTTCAACTTCCTTTTTACCTTAAAGGCCCGGGGATGGACTTGCAGTGGGCCCC 540

Qy   2087 atgtagttttaga 2099
          |||||||||||||
Db    541 ATGTAGTTTTAGA 553


RESULT   5
AW660536
LOCUS       AW660536      552 bp    mRNA            EST     14-JUL-2000
DEFINITION  99459 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION   AW660536
VERSION     AW660536.1  GI:7426363
KEYWORDS    EST.
SOURCE      cow.
  ORGANISM  Bos taurus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
            Bovidae; Bovinae; Bos.
  REFERENCE 1  (bases 1 to 552)
  AUTHORS   Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M.,
            Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid
            ,W.W. and Keele,J.W.
  TITLE     Design and use of four pooled tissue normalized cDNA libraries for
            EST discovery in cattle
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR PRimers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 90  row: O  column: 10
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1. .552
                     /organism="Bos taurus"
                     /db_xref="taxon:9913"
                     /clone_lib="MARC 1BOV"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
                     Library made from pooled tissue from lymph node, ovary,
                     fat, hypothalamus, and pituitary."
BASE COUNT      146 a    113 c    152 g    141 t
ORIGIN


  Query Match           19.2%;  Score 479;  DB 119;  Length 552;
  Best Local Similarity   91.8%;  Pred. No. 4.6e-124;
  Matches  506;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy   1104 gctccatgagttctcaaaatacggcagcgcattaagcagcatgctgatgctgtgcacct 1163
          |||||| || ||||| ||||| || ||||| || |  ||||||||||||||||||||| ||
Db      2 GCTCCACGAATTCTCCAAATATCGACAGCGTATCAGACAGCATGCTGATGCTGTGCATCT 61

Qy   1164 catctcgcgggtgacatttcactataagagaagcagtctgagttactttggaggtgtctg 1223
          ||| ||||| |||||||| |||||||||| ||||| ||||||||||||||||||||||
Db     62 CATTTCGCGTGTGACATTCCACTATAAGAGGAGCAGCCTGAGTTACTTTGGAGGTGTCTG 121

Qy   1224 ttctcgcacaagaggagttggtgtgaatgagtatggtcttccaatggcagtggcacaagt 1283
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    122 TTCTCGCACAAGGGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACAAGT 181

Qy   1284 attatcgcagagcctggctcaaaaccttggaatccaatgggaaccttctagcagaaagcc 1343
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    182 ATTATCGCAGAGCCTGGCTCAGAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCC 241

Qy   1344 aaaatgtgactgcacagaatcctggggtggctgcatcatggaggaaacaggggtgtccca 1403
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||||| || ||
Db    242 AAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCATGGAAGAAACAGGGGTATCTCA 301

Qy   1404 ttctcgaaaattttcaaagtgcagcattttggagtatagagacttttacagagaggagg 1463
          |||  |||| || |||||||||||| |||||||| |||||||||||||||| |||||| ||
Db    302 TTCCAGAAAGTTCTCAAAGTGCAGTATTTTGGAATATAGAGACTTTTTACTGAGAGGGGG 361

Qy   1464 tggagcctgccttttcaacaggccaacaaagctatttgagcccacggaatgtggaaatgg 1523
          ||||||||| |||||||||||||||||||||||||||||||||||| ||||||||||||||
Db    362 TGGAGCCTGTCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACAGAATGTGGAAATGG 421

Qy   1524 atacgtggaagctggggaggagtgtgattgtggttttcatgtggaatgctatggattatg 1583
          |||||| |||||||| |||||||||||||||||| | |||||||||||||||||| | ||
Db    422 ATACGTAGAAGCTGGAGAGGAGTGTGATTGTGGTATCCATGTGGAATGCTATGGACTGTG 481

Qy   1584 ctgtaagaaatgttccctctccaacggggctcactgcagcgacgggccctgctgtaacaa 1643
          |||||||||||| || |||||||| ||||| |||||||||||| ||||| ||||| || ||
Db    482 CTGTAAGAAATGCTCTCTCTCCAATGGGGCCCACTGCAGCGATGGGCCTTGCTGCAATAA 541

Qy   1644 tacctcatgtc 1654
          |||||| ||||
Db    542 TACCTCGTGTC 552


RESULT   6
BF905823
LOCUS       BF905823      447 bp    mRNA            EST     18-JAN-2001
DEFINITION  IL3-MT0267-271200-411-A09 MT0267 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF905823
VERSION     BF905823.1  GI:12297282
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 447)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=IL3&t2=IL3-MT0267-
            271200-411-A09&t3=2000-12-27&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 10
            High quality sequence stop: 66.
FEATURES             Location/Qualifiers
     source          1. .447
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="MT0267"
                     /dev_stage="Adult"
```

/note="Organ: marrow; Vector: puc18, Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 122 a 104 c 120 g 98 t 3 others
ORIGIN

Query Match 16.5%; Score 413.2; DB 171; Length 447;
Best Local Similarity 95.3%; Pred. No. 1.7e-105;
Matches 424; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

```
Qy   1583 gctgtaagaaatgttccctctccaacggggctcactgcagcgacgggccctgctgtaaca 1642
          |||   |   || |||||||||  ||||||||||||||||||||||||||||||||||||
Db      3 GCTCCGACCCATCTTCCCTCTCTCACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACA 62

Qy   1643 atacctcatgtctttttcagccacgagggtatgaatgccgggatgctgtgaacgagtgtg 1702
          ||||| || | | ||||||||||||||||||||||||||||||||||||||||||||||
Db     63 ATACCACANGNCNTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTG 122

Qy   1703 atattactgaatattgtactggagactctggtcagtgcccaccaaatcttcataagcaag 1762
          |||||||||||||||||||||||||   |  || ||||||||||||| ||| |||||
Db    123 ATATTACTGAATATTGTACTGGAGACTACTGGGAGGGCCCACCAAATCTTTATATGCAAG 182

Qy   1763 acggatatgcatgcaatcaaaatcagggccgctgctacaatggcgagtgcaagaccagag 1822
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    183 ACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGAG 242

Qy   1823 acaaccagtgtcagtacatctggggaacaaaggctgcagggtctgacaagttctgctatg 1882
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    243 ACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATG 302

Qy   1883 aaaagctgaatacagaaggcactgagaagggaaactgcgggaaggatggagaccggtgga 1942
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    303 AAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGA 362

Qy   1943 ttcagtgcagcaaacatgatgtgttctgtggattcttactctgtaccaatcttactcgag 2002
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    363 TTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAG 422

Qy   2003 ctccacgtattggtcaacttcaggg 2027
          |||||||||||||||||||||||||
Db    423 CTCCACGTATTGGTCAACTTCAGGG 447
```

RESULT 7
BF308417
LOCUS BF308417 503 bp mRNA EST 21-NOV-2000
DEFINITION 601890734F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4131777 5', mRNA sequence.
ACCESSION BF308417
VERSION BF308417.1 GI:11255636
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 503)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM1030 row: c column: 10
High quality sequence stop: 503.
FEATURES Location/Qualifiers
source 1. .503
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4131777"
/clone_lib="NIH_MGC_17"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: muscle; Vector: pOTB7; Site_1: EcoRI; Site_2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 130 a 103 c 142 g 128 t
ORIGIN

Query Match 15.9%; Score 396.4; DB 147; Length 503;
Best Local Similarity 96.7%; Pred. No. 1e-100;
Matches 491; Conservative 0; Mismatches 6; Indels 11; Gaps 8;

```
Qy   1161 cctcatctcgcgggtgacatttcactataagagaagcagtctgagttactttggaggtgt 1220
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db      2 CCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTAC-TTGGAGGTGT 60

Qy   1221 ctgttctcgcacaagaggagttggtgtgaatgagtatggtcttccaatggcagtggcaca 1280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACA 120

Qy   1281 agtattatcgcagagcctggctcaaaaccttggaatccaa-tgggaaccttctagcagaa 1339
          |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db    121 AGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAAGTGGGAACCTTCTAGCAGAA 180

Qy   1340 agccaaaatgtgactgcacagaatcctggggtggctgca-tcatggaggaaacaggggtg 1398
          |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db    181 AGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCACTCATGGAGGAAACAGGGGTG 240

Qy   1399 tcccattctcgaaaattttcaaagtgcagcattttggagtatagagacttttttacagaga 1458
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    241 TCCCATTCTCGAAAATTTTCAAAGTGCAGCA-TTTGGAGTATAGAGACTTTTTACAGAGA 299

Qy   1459 ggaggtggagcctgccttttcaacaggccaacaaagctatttgagcccacggaatgtgg- 1517
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    300 GGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGA 359

Qy   1518 -aaatggatacgtggaagctggggaggagtgtgattgtggttttca-tgtggaatgctat 1575
           | ||||||||||||||||||||||||||||||| |||||||||| ||||||||||||||
Db    360 CACATGGATACGTGGAAGCTGGGGAGGAGTGTGATCGTGGTTTTCACTGTGGAATGCTAT 419

Qy   1576 ggattatgctgtaagaaatgttccctctccaacggggctcactgcagcgacgggccctgc 1635
          |||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    420 GGATTATGCTGTAAGAAATGTTCCCTCTCCAACGGGGCTCACTG-AGCGACGGGCCTGTG 478

Qy   1636 tgtaacaatacctcatgtctttttcagc 1663
          |   |||||||||||||||||||||||||
Db    479 T---ACAATACCTCATGTCTTTTTCAGC 503
```

RESULT 8
AA718688
LOCUS AA718688 428 bp mRNA EST 29-DEC-1997
DEFINITION vt87c03.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone IMAGE:1178116 5' similar to TR:Q14809 Q14809 METALLOPROTEASE/DISINTEGRIN-LIKE PROTEIN ;, mRNA sequence.
ACCESSION AA718688
VERSION AA718688.1 GI:2730962
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 428)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:635964
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 391.
     RES             Location/Qualifiers
     source          1. .428
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1178116"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
                     ) with a modified polylinker; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT      117 a    107 c    114 g     90 t
ORIGIN


  Query Match           13.5%;  Score 336.8;  DB 11;  Length 428;
  Best Local Similarity  86.7%;  Pred. No. 6.7e-84;
  Matches  371;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

      617 ctgcagcgacgggccctgctgtaacaatacctcatgtctttttcagccacgagggtatga 1676
          |||||| ||||  |||||||||||||| |||||||||||||||||| ||||||||||||
        1 CTGCAGTGACGCCCCCTGCTGTAACAACACCTCATGTCTTTTTCAGTCACGAGGGTATGA 60

Qy   1677 atgccgggatgctgtgaacgagtgtgatattactgaatattgtactggagactctggtca 1736
          ||| |||||||| || |||   |||||||| || || || || ||||||||||||||| ||
Db     61 ATGTCGGGATGCCGTAAACAGCTGTGATATCACCGAGTACTGCACTGGAGACTCTGGCCA 120

Qy   1737 gtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctg 1796
          ||||||||| || || ||||| ||||| || |||   |||||||||||||||| |||||
Db    121 GTGCCCACCGAACCTCCATAAACAAGATGGCTATAGCTGCAATCAAAATCAGGGTCGCTG 180

Qy   1797 ctacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctggggaacaaaggc 1856
          |||||||||||||||||||| || ||||| || || ||||||||||||| ||||||||
Db    181 CTACAATGGCGAGTGCAAGACAAGGGACAATCAATGCCAGTACATCTGGGGGACAAAGGC 240

Qy   1857 tgcagggtctgacaagttctgctatgaaaagctgaatacagaaggcactgagaagggaaa 1916
          ||| ||||| |||||||||||||||||||||||||| || |||||||| || ||||| ||
Db    241 TGCGGGGTCAGACAAGTTCTGCTATGAAAAGCTGAACACGGAAGGCACCGACAAGGGCAA 300

Qy   1917 ctgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtggatt 1976
          || || |||||||||||||||||| | |||||||| ||||||||||||||||||||
Db    301 TTGTGGAAAGGATGGAGACCGGTGGATCCCGTGCAGCAAGCATGATGTGTTCTGTGGATT 360

Qy   1977 cttactctgtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatcat 2036
          | || || |||||||||| |||||||||||||| |||||||||| || ||||||||
Db    361 TCTGCTTTGCACCAATCTTACCCGAGCTCCACGTATCGGTCAACTTCAAGGAGAGATCAT 420

Qy   2037 tccaactt 2044
          || ||||
Db    421 CCCGACTT 428


RESULT   9
F08148
LOCUS       F08148         340 bp    mRNA            EST       21-FEB-1995
DEFINITION  HSC2RF011 normalized infant brain cDNA Homo sapiens cDNA clone
            c-2rf01, mRNA sequence.
ACCESSION   F08148
VERSION     F08148.1  GI:677656
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 340)
  AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
            ,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
            Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
            Sebastiani-Kabaktchis,C. and Tessier,A.
  TITLE     IMAGE: molecular integration of the analysis of the human genome
            and its expression
  JOURNAL   C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
  MEDLINE   95277534
COMMENT     Contact: Genethon
            Genexpress-Genethon
            Genethon Centre de recherche sur le Genome Humain
            1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
            Tel: 33169472800
            Fax: 33160778698
            Email: genexpress@genethon.fr
            Single read.
            Genexpress_library_idt: C; Genexpress_sequence_idt: ylc-2rf01
            Seq primer: (-21)M13_universal.
FEATURES             Location/Qualifiers
     source          1. .340
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="c-2rf01"
                     /clone_lib="normalized infant brain cDNA"
                     /sex="Female"
                     /tissue_type="total brain"
                     /dev_stage="3 months old"
                     /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
                     Site_2: NotI; sex=Female; dev_stage=3 months old;
                     isolate=muscular atrophy patient;  tissue_type=total brain
                     ;  total mRNA was oligo-(dT) primed and directionally
                     cloned 5' -> 3' into the  HindIII -> NotI sites of the
                     lafmid BA vector. Clone library from B.Soares,  Psychiatry
                     Dept. Columbia University, USA.  Normalization_method:
                     Bento Soares, P.N.A.S in press"
BASE COUNT       87 a     70 c     90 g     84 t      9 others
ORIGIN


  Query Match           13.2%;  Score 329.4;  DB 157;  Length 340;
  Best Local Similarity  97.1%;  Pred. No. 7.6e-82;
  Matches  330;  Conservative    0;  Mismatches   10;  Indels    0;  Gaps    0;

Qy   1463 gtggagcctgccttttcaacaggccaacaaagctatttgagcccacggaatgtggaaatg 1522
          ||||||||||||||| |||||||  |||||||||||| ||||||||||||||||||||
Db      1 GTGGAGCCTGCCTTTCCAACAGGNNAACAAAGCTATTTNAGCCCACGGAATGTGGAAATG 60

Qy   1523 gatacgtggaagctggggaggagtgtgattgtggttttcatgtggaatgctatggattat 1582
```

```
Db     61 GNTNCGTGGAAGCTGGGGAGGAGTGTNATTGTGGTTTTNATGTGGAATGCTATGGATTAT 120

Qy   1583 gctgtaagaaatgttccctctccaacgggggctcactgcagcgacgggccctgctgtaaca 1642
          |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 NCTGTAAGAAATGTNCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACA 180

Qy   1643 atacctcatgtctttttcagccacgagggtatgaatgccgggatgctgtgaacgagtgtg 1702
          |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    181 ATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTG 240

Qy   1703 atattactgaatattgtactggagactctggtcagtgcccaccaaatcttcataagcaag 1762
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAG 300

Qy   1763 acggatatgcatgcaatcaaaatcagggccgctgctacaa 1802
          |||||||||||||||||||||||||||||||||||||||
Db    301 ACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAA 340


RESULT  10
AA511039
LOCUS       AA511039      375 bp    mRNA            EST       08-JUL-1997
DEFINITION  vh55g05.r1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone
            IMAGE:890936 5' similar to TR:G836683 G836683
            METALLOPROTEASE/DISINTEGRIN-LIKE PROTEIN ;, mRNA sequence.
ACCESSION   AA511039
VERSION     AA511039.1  GI:2248893
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 375)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:518896
            Seq primer: -28m13 rev2 ET from Amersham
            High quality sequence stop: 312.
FEATURES             Location/Qualifiers
     source          1. .375
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:890936"
                     /clone_lib="Soares_mammary_gland_NbMMG"
                     /sex="male"
                     /tissue_type="mammary gland"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: mammary gland; Vector: pT7T3D-Pac (Pharmacia
                     ) with a modified polylinker; Site_1: Not I; Site_2: Eco
                     RI; 1st strand cDNA was primed with a Not I - oligo(dT)
                     primer [5'
                     TGTTACCAATCTGAAGTGGGAGCGGCCGCGAATGGTTTTTTTTTTTTTTTTTTTTTT
                     T 3']; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
                     constructed and normalized by Bento Soares and M.Fatima
                     Bonaldo."
BASE COUNT      106 a     86 c    107 g     76 t
ORIGIN


  Query Match           11.8%;  Score 295;  DB 8;  Length 375;
  Best Local Similarity  86.7%;  Pred. No. 4.2e-72;
  Matches  325;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy   1657 tttcagccacgagggtatgaatgccgggatgctgtgaacgagtgtgatattactgaatat 1716
          |||||| |||||||||||||||| |||||||| || |||   |||||||| || || ||
Db      1 TTTCAGTCACGAGGGTATGAATGTCGGGATGCCGTAAACAGCTGTGATATCACCGAGTAC 60

Qy   1717 tgtactggagactctggtcagtgcccaccaaatcttcataagcaagacggatatgcatgc 1776
          || |||||||||||||| |||||||||| || || ||||| ||||| || |||   |||
Db       61 TGCACTGGAGACTCTGGCCAGTGCCCACCGAACCTCCATAAACAAGATGGCTATAGCTGC 120

Qy   1777 aatcaaaatcagggccgctgctacaatggcgagtgcaagaccagagacaaccagtgtcag 1836
          |||||||||||||| ||||||||||||||||||||||||||| || ||||| || || |||
Db     121 AATCAAAATCAGGGTCGCTGCTACAATGGCGAGTGCAAGACAAGGGACAATCAATGCCAG 180

Qy   1837 tacatctggggaacaaaggctgcagggtctgacaagttctgctatgaaaagctgaataca 1896
          |||||||||| |||||||||| ||||| ||||||||||||||||||||||||||| ||
Db     181 TACATCTGGGGGACAAAGGCTGCGGGGTCAGACAAGTTCTGCTATGAAAAGCTGAACACG 240

Qy   1897 gaaggcactgagaagggaaactgcgggaaggatggagaccggtggattcagtgcagcaaa 1956
          |||||||| |||||||| || || || |||||||||||||||||| | |||||||||
Db     241 GAAGGCACCGAGAAGGGCAATTGTGGAAAGGATGGAGACCGGTGGATCCCGTGCAGCAAG 300

Qy   1957 catgatgtgttctgtggattcttactctgtaccaatcttactcgagctccacgtattggt 2016
          |||||||||||||||||||  | || || |||||||||| ||||||||||||||| |||
Db     301 CATGATGTGTTCTGTGGATTTCTGCTTTGCACCAATCTTACCCGAGCTCCACGTATCGGT 360

Qy   2017 caacttcagggtgag 2031
          |||||||| || |||
Db     361 CAACTTCAAGGAGAG 375


RESULT  11
R52569
LOCUS       R52569        405 bp    mRNA            EST       18-MAY-1995
DEFINITION  yg81c01.r1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:39972 5' similar to SP:DISI_BITAR P17497 DISINTEGRIN BITAN ;,
            mRNA sequence.
ACCESSION   R52569
VERSION     R52569.1  GI:814471
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 405)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1295
            High quality sequence stops: 53 Source: IMAGE Consortium, LLNL This
            clone is available royalty-free through LLNL ; contact the IMAGE
            Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1295   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 53.
FEATURES             Location/Qualifiers
```

```
     source          1. .405
                     /organism="Homo sapiens"
                     /db_xref="GDB:412513"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:39972"
                     /clone_lib="Soares infant brain 1NIB"
                     /sex="female"
                     /dev_stage="73 days post natal"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                     I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                     I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Hind III adaptors
                     (Pharmacia), digested with Not I and directionally cloned
                     into the Not I and Hind III sites of the Lafmid BA vector.
                     Library went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      105 a     84 c    111 g     97 t      8 others
ORIGIN


[redacted]ry Match             10.9%;  Score 271.2;  DB 187;  Length 405;
[redacted]st Local Similarity   88.4%;  Pred. No. 2.3e-65;
  Matches  351;  Conservative    0;  Mismatches   33;  Indels   13;  Gaps    5;

Qy   1464 tggagcctgccttttcaacaggccaacaaagctatttgagcccacggaatgtggaaatgg 1523
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGG 60

Qy   1524 atacgtggaagctggggaggagtgtgattgtggttttcatgtggaatgctatggattatg 1583
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTATGGATTATG 120

Qy   1584 ctgtaagaaatgttccctctccaacggggctcactgcagcgacgggccctgctgtaacaa 1643
           |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    121 CTGTAAGAAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGNAACAA 180

Qy   1644 tacctcatgtctttttcagccacgagggtatgaatgccgggatgctgtgaacgagtgtga 1703
           |||||||||||||||||||||||| |||||||||||| |||||||||||||||||| |
Db    181 TACCTCATGTCTTTTTCAGCCACGANGGTATGAATGCCNGGATGCTGTGAACGAGTGTTA 240

Qy   1704 tattactgaata-ttgtactggagactctggtcagtgcccaccaaatcttcataagcaag 1762
           |||||||||||| |||||||||||||||||||||||||||||||||  |   |
Db    241 TATTACTGAATATTTGTACTGGAGACTCTGGTCAGTGCCCACCAAATTCTTTCATAAGGC 300

Qy   1763 acggatatgcatgcaatcaaaatca--------gggccgctgct-acaatggcgag-tgc 1812
           | |       |||  ||   |||||        |||||||||| |||||||||| |||
Db    301 AAGACGGGATATGGCATGGCAATCAAAATNCAGGGGCCGCTGCTNACAATGGCGAGTTGC 360

Qy   1813 aagaccagagacaa--ccagtgtcagtacatctgggg 1847
           |||||||| |   |  |||||||||||||||| |||
Db    361 AAGACCAGNGGACANCCCAGTGTCAGTACATCTNGGG 397


RESULT  12
R15038
LOCUS       R15038          326 bp    mRNA            EST       13-APR-1995
DEFINITION  yf86d08.r1 Soares infant brain 1NIB Homo sapiens cDNA clone
            IMAGE:29473 5' similar to SP:S23403 S23403; SPERM SURFACE PROTEIN
            PH-30 BETA CHAIN PRECURSOR - GUINEA PIG ;, mRNA sequence.
ACCESSION   R15038
VERSION     R15038.1  GI:769311
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 326)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Insert Size: 1226
            High quality sequence stops: 278 Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1226   Std Error: 0.00
            Seq primer: M13RP1
            High quality sequence stop: 278.
FEATURES             Location/Qualifiers
     source          1. .326
                     /organism="Homo sapiens"
                     /db_xref="GDB:401820"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:29473"
                     /clone_lib="Soares infant brain 1NIB"
                     /sex="female"
                     /dev_stage="73 days post natal"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
                     I; Site_2: Hind III; 1st strand cDNA was primed with a Not
                     I - oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCAGGAATTTTTTTTTTTTTTTTT 3'];
                     double-stranded cDNA was ligated to Hind III adaptors
                     (Pharmacia), digested with Not I and directionally cloned
                     into the Not I and Hind III sites of the Lafmid BA vector.
                     Library went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT       82 a     67 c     91 g     86 t
ORIGIN


  Query Match             10.7%;  Score 267;  DB 187;  Length 326;
  Best Local Similarity   99.3%;  Pred. No. 3.3e-64;
  Matches  289;  Conservative    0;  Mismatches    0;  Indels    2;  Gaps    2;

Qy   1464 tggagcctgccttttcaacaggccaacaaagctatttgagcccacggaatgtggaaatgg 1523
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGG 60

Qy   1524 atacgtggaagctggggaggagtgtgattgtggttttcatgtggaatgctatggattatg 1583
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTATGGATTATG 120

Qy   1584 ctgtaagaaatgttccctctccaacggggctcactgcagcgacgggccctgctgtaacaa 1643
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CTGTAAGAAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAA 180

Qy   1644 tacctcatgtctttttcagccacgagggtatgaatgccgggatgctgtgaacgagtgtga 1703
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGA 240

Qy   1704 tattac-tgaatattgtact-ggagactctggtcagtgcccaccaaatctt 1752
           |||||| |||||||||||||| |||||||||||||||||||||||||||||
Db    241 TATTACTTGAATATTGTACTGGGAGACTCTGGTCAGTGCCCACCAAATCTT 291


RESULT  13
BF693020
LOCUS       BF693020        763 bp    mRNA            EST       22-DEC-2000
DEFINITION  602080016F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:4244595 5',
            mRNA sequence.
ACCESSION   BF693020
VERSION     BF693020.1  GI:11978428
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
```

```
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 763)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONETECH Laboratories, Inc.
            cDNA Library Preparation: CLONETECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1056  row: p  column: 04
            High quality sequence stop: 639.
FEATURES             Location/Qualifiers
     source          1. .763
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4244595"
                     /clone_lib="NIH_MGC_81"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: muscle (skeletal); Vector: pDNR-LIB
                     (Clontech); Site_1: SfiI (ggccgcctcgggcc); Site_2: SfiI
                     (ggccattatggcc); 5' and 3' adaptors were used in cloning
                     as follows: 5' adaptor sequence:  5'-CACGGCCATTATGGCC-3'
                     and 3' adaptor sequence:
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)BN-3' (where B = A,
                     C, or G and N = A, C,  G, or T).  Average insert size
                     1.55 kb (range 1.0-4.0 kb).  15/15 colonies contained
                     inserts by PCR. This library was enriched for full-length
                     clones and was constructed by Clontech  Laboratories (Palo
                     Alto, CA)."
BASE COUNT      178 a    169 c    193 g    221 t      2 others
ORIGIN


  Query Match             10.4%;  Score 259.8;  DB 168;  Length 763;
  Best Local Similarity   96.8%;  Pred. No. 4.7e-62;
  Matches  275;  Conservative    0;  Mismatches    8;  Indels    1;  Gaps    1;

Qy   1367 ggggtggctgcatcatggaggaaacaggggtgtcccattctcgaaaattttcaaagtgca 1426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCA 60

Qy   1427 gcattttggagtatagagacttttttacagagaggaggtggagcctgcctttttcaacaggc 1486
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GCATTTGGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGC 120

Qy   1487 caacaaagctatt-tgagcccacggaatgtggaaatggatacgtggaagctggggaggag 1545
           |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    121 CAACAAAGCTATTGTGAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAG 180

Qy   1546 tgtgattgtggttttcatgtggaatgctatggattatgctgtaagaaatgttccctctcc 1605
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGTGATGGTGGTTTTCATGTGGAATGCTATGGATTATGCTGTAAGAAATGTTCCCTCTCC 240

Qy   1606 aacggggctcactgcagcgacgggccctgctgtaacaatacctc 1649
           |||||||||||||||||||||||||||  |||||||    |||
Db    241 AACGGGGCTCACTGCAGCGACGGGCCCTNGTGTAACATCCTCTC 284


RESULT  14
AA317222
LOCUS       AA317222       261 bp    mRNA            EST       19-APR-1997
DEFINITION  EST19173 Retina II Homo sapiens cDNA 5' end similar to similar to
            metalloproteinase, hemorrhagic, mRNA sequence.
ACCESSION   AA317222
VERSION     AA317222.1  GI:1969602
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 261)
  AUTHORS   Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult
            ,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White
            ,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A.,
            Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald
            ,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A.,
            Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M.,
            Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
            Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
            Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
            Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
            Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
            Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W.
            , Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L.,
            Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L.,
            Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon
            ,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and
            Venter,J.C.
  TITLE     Initial assessment of human gene diversity and expression patterns
            based upon 83 million nucleotides of cDNA sequence
  JOURNAL   Nature 377 (6547 Suppl), 3-174 (1995)
  MEDLINE   96026280
COMMENT     Other_ESTs: THC117909
            Contact: Kerlavage, AR
            Bioinformatics
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850 USA
            Tel: 3018699056
            Fax: 3018699423
            Email: arkerlav@tigr.org
            For clone availability, additional sequence and expression
            information related to this EST, please check the TIGR Human Gene
            Index (http://www.tigr.org/tdb/hgi/hgi.html)
            Seq primer: M13 Reverse.
FEATURES             Location/Qualifiers
     source          1. .261
                     /organism="Homo sapiens"
                     /db_xref="ATCC (inhost):117726"
                     /db_xref="taxon:9606"
                     /clone_lib="Retina II"
                     /sex="mixed"
                     /dev_stage="adult"
                     /note="Organ: retina; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
BASE COUNT       63 a     56 c     68 g     70 t      4 others
ORIGIN


  Query Match             10.3%;  Score 257;  DB 5;  Length 261;
  Best Local Similarity   98.5%;  Pred. No. 2e-61;
  Matches  257;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy   1547 gtgattgtggttttcatgtggaatgctatggattatgctgtaagaaatgttccctctcca 1606
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GTGATTGTGGTTTTCATGTGGAATGCTATGGATTATGCTGTAAGAAATGTTCCCTCTCCA 60

Qy   1607 acggggctcactgcagcgacgggccctgctgtaacaatacctcatgtctttttcagccac 1666
           |||||||||||||||||||||||||||||||||| |||||||||| ||||||||||||
Db     61 ACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACANTACCTCATGTNTTTTTCAGCCAC 120

Qy   1667 gagggtatgaatgccgggatgctgtgaacgagtgtgatattactgaatattgtactggag 1726
           |||||||||| |||||||||||||||||||||||||||||||| ||||||||||||||
Db    121 GAGGGTATGANTGCCGGGATGCTGTGAACGAGTGTGATATTACTGANTATTGTACTGGAG 180

Qy   1727 actctggtcagtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatc 1786
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATC 240

Qy   1787 agggccgctgctacaatggcg 1807
           |||||||||||||||||||||
```

```
Db    241 AGGGCCGCTGCTACAATGGCG 261


RESULT 15
AW898896/c
LOCUS       AW898896      397 bp    mRNA            EST       24-MAY-2000
DEFINITION  CM0-NN0079-140400-334-g06 NN0079 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW898896
VERSION     AW898896.1  GI:8063101
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 397)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=CM0-NN0079-140
            400-334-g06&t3=2000-04-14&t4=1)
            Seq primer: puc 18 forward
            High quality sequence stop: 397.
FEATURES             Location/Qualifiers
     source          1. .397
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NN0079"
                     /dev_stage="Adult"
                     /note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT      110 a    114 c     79 g     94 t


  Query Match             9.3%;  Score 232;  DB 122;  Length 397;
  Best Local Similarity   92.9%;  Pred. No. 2.7e-54;
  Matches  299;  Conservative    0;  Mismatches   15;  Indels    8;  Gaps    5;

Qy   1421 agtgcagcattttggagtatagagac-tttttacagagaggaggtggagcctgccttttc 1479
            ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db    319 AGTGCAGCATTTTGGAGTATAGAGACTTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTC 260

Qy   1480 aacaggccaacaaagctat-ttgagcccacggaatgtggaaatggatacgtggaagctgg 1538
            |||||||||| ||||||| ||||||||||||||||||||||||||||||||||||||
Db    259 AACAGGCCAACTAAGCTATCTTGAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGG 200

Qy   1539 ggaggagtgtgattgtggttttcatgtggaatgctatggattatgctgtaagaaatgttc 1598
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    199 GGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTATGGATTATGCTGTAAGAAATGTTC 140

Qy   1599 cctctccaacggggctcactgcagcgacggg--ccctgctgtaacaatacctcatgtctt 1656
            |||||||||||||||||||||||||||||  ||   |||| ||||||||||||||||
Db    139 CCTCTCCAACGGGGCTCACTGCAGCGACGGGACCCTGTCTGTCACAATACCTCATGTCTT 80

Qy   1657 tttcagcc-acgagggtatgaatgccgggatgctgtgaacgagtgtgatattactgaata 1715
            |||||||| |||||||||||||||||||||||||||||    | |||  ||  |||||||
Db     79 TTTCAGCCAACGAGGGTATGAATGCCGGGATGCTGTGA---CGAGTGTGATATCTGAATA 23

Qy   1716 ttgtactggagactctggtcag 1737
            |||||||||||||   ||| |
Db     22 TTGTACTGGAGACTAACGTCCG 1


RESULT 16
AW900312/c
LOCUS       AW900312      237 bp    mRNA            EST       24-MAY-2000
DEFINITION  CM4-NN1004-010400-129-c04 NN1004 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW900312
VERSION     AW900312.1  GI:8064517
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 237)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
            ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=CM4-NN1004-010
            400-129-c04&t3=2000-04-01&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 6
            High quality sequence stop: 237.
FEATURES             Location/Qualifiers
     source          1. .237
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone_lib="NN1004"
                     /dev_stage="Adult"
                     /note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from  ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 -  Ludwig Institute for Cancer Research)
                     profiles into  the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
BASE COUNT       65 a     66 c     49 g     57 t
ORIGIN


  Query Match             9.2%;  Score 228.8;  DB 122;  Length 237;
  Best Local Similarity   99.1%;  Pred. No. 1.9e-53;
  Matches  230;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy   2029 gagatcattccaacttccttctaccatcaaggccgggtgattgactgcagtggtgcccat 2088
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    237 GAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCAT 178
```

```
Qy   2089 gtagttttagatgatgatacggatgtgggctatgtagaagatggaacgccatgtggcccg 2148
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    177 GTAGTTTTAGATGATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCG 118

Qy   2149 tctatgatgtgtttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgt 2208
          |||||||||||||||||||||| ||||||| ||||||||||||||||||||||||||||
Db    117 TCTATGATGTGTTTAGATCGGAAGCGCCTACATATTCAAGCCCTAAATATGAGCAGCTGT 58

Qy   2209 ccactcgattccaagggtaaagtctgttcgggccatggggtgtgtagtaatg 2260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     57 CCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTACTAATG 6
```

LT 17
396/c

```
            AI342396     576 bp    mRNA            EST       13-FEB-1999
NITION      qt27g06.x1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone
            IMAGE:1949242 3' similar to TR:Q14809 Q14809
            METALLOPROTEASE/DISINTEGRIN-LIKE PROTEIN ;, mRNA sequence.
ACCESSION   AI342396
VERSION     AI342396.1  GI:4079323
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 576)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 1212   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 418.
FEATURES             Location/Qualifiers
     source          1. .576
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1949242"
                     /clone_lib="Soares_pregnant_uterus_NbHPU"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
                     Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
                     oligo(dT) primer [5'
                     AACTGGAAGAATTCGCGGCCGCCTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Not I and cloned into the Not I
                     and Eco RI sites of the modified pT7T3 vector. Library
                     went through one round of normalization.  Library
                     constructed by M. Fatima Bonaldo."
BASE COUNT      133 a    139 c    128 g    175 t      1 others
ORIGIN
```

```
  Query Match             8.4%;  Score 211;  DB 19;  Length 576;
  Best Local Similarity   97.4%;  Pred. No. 2.7e-48;
  Matches  225;  Conservative    0;  Mismatches    5;  Indels    1;  Gaps    1;

Qy   2269 tgcatttgtgatttcacctgggcagggacagattgcagtatccgggatccagttaggaac 2328
          ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||  ||
Db    576 TGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTATCGGGGATCCAGTTAGG-AC 518

Qy   2329 cttcacccccccaaggatgaaggacccaagggtcctagtgccaccaatctcataataggc 2388
          ||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    517 CTCACCCCCCCCAAGGATGAAGGACCCAAGGGTCCTAGTGCCACCAATCTCATAATAGGC 458
```

```
Qy   2389 tccatcgctggtgccatcctggtagcagctattgtccttgggggcacaggctgggattt 2448
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    457 TCCATCGCTGGTGCCATCCTGGTAGCAGCTATTGTCCTTGGGGGCACAGGCTGGGGATTT 398

Qy   2449 aaaaatgtcaagaagagaaggttcgatcctactcagcaaggccccatctga 2499
          |||||||||||||||||||||||||||||||||||||||||||||| |||
Db    397 AAAAATGTCAAGAAGAGAAGGTTCGATCCTACTCAGCAAGGCCCCATTTGA 347
```

```
RESULT  18
BE309477
LOCUS       BE309477      504 bp    mRNA            EST       26-OCT-2000
DEFINITION  601095217F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:3489677 5',
            mRNA sequence.
ACCESSION   BE309477
VERSION     BE309477.1  GI:9167726
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 504)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM8531  row: a  column: 06
            High quality sequence stop: 478.
FEATURES             Location/Qualifiers
     source          1. .504
                     /organism="Mus musculus"
                     /strain="C57/B6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3489677"
                     /clone_lib="NCI_CGAP_Mam5"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="7 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Robin Humphreys,
                     NIH"
BASE COUNT      130 a    113 c    143 g    118 t
ORIGIN
```

```
  Query Match             7.6%;  Score 190.6;  DB 166;  Length 504;
  Best Local Similarity   84.5%;  Pred. No. 1.5e-42;
  Matches  273;  Conservative    0;  Mismatches   44;  Indels    6;  Gaps    5;

Qy   1831 tgtcagtacatctggggaacaaaggctgc-agggtctgacaagttctgctatgaaaagct 1889
          || |||||||||||||| ||||||||||  ||||| ||||||||||||||||||||||||
Db      2 TGCCAGTACATCTGGGGGACAAAGGCTGCGTGGGTCAGACAAGTTCTGCTATGAAAAGCT 61

Qy   1890 gaatacagaaggcactgagaaggg-aaactgcgggaaggatggagaccggtggattcagt 1948
          ||| || |||||||| |||||||| ||  || || ||||||||||||||||||| | ||
Db     62 GAACACGGAAGGCACCGAGAAGGGCAATATGTGGAAAGGATGGAGACCGGTGGATCCCGT 121

Qy   1949 gcagcaa-acatgatgtgttctgtggattcttactctgtaccaatcttactcgagctcca 2007
          |||||||  |||||||||||||||||||  | ||  | ||||||||||| |||||||||
Db    122 GCAGCAATGCATGATGTGTTCTGTGGATTTCTGCTTAGCACCAATCTTACCCGAGCTCCA 181

Qy   2008 cg-tattggtcaacttcagggtgagatcattccaacttccttctaccatcaaggccgggt 2066
          || ||| |||||||||||| || |||||||| || |||||||||||| |||||||||| ||
```

```
Db    182 CGATATCGGTCAACTTCAAGGAGAGATCATCCCGACTTCCTTCTATCATCAAGGCCGAGT 241

Qy   2067 gattgactgcagtggtgcccatgtagt--tttagatgatgatacggatgtgggctatgta 2124
          ||||||||||||||||||| ||||||||  ||    |||||||||| || ||||| || |||
Db    242 GATTGACTGCAGTGGTGCTCATGTAGTTATTAGACTGATGATACAGACGTGGGTTACGTA 301

Qy   2125 gaagatggaacgccatgtggccc 2147
          |||||||| || ||  |  | ||
Db    302 GAAGATGGGACTCCGAGATGGCC 324


RESULT  19
AI597498
LOCUS       AI597498      476 bp    mRNA            EST       21-APR-1999
DEFINITION  vj50a07.x1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
            IMAGE:932436 3' similar to TR:Q14809 Q14809
            METALLOPROTEASE/DISINTEGRIN-LIKE PROTEIN ;, mRNA sequence.
ACCESSION   AI597498
VERSION     AI597498.1  GI:4606546
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 476)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:537356
            This clone was previously sequenced on the 5' end only, this new
            data is from the 3' end
            Possible reversed clone: similarity on wrong strand
            High quality sequence stop: 398.
FEATURES             Location/Qualifiers
     source          1. .476
                     /organism="Mus musculus"
                     /strain="NIH/Swiss"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:932436"
                     /clone_lib="Stratagene mouse heart (#937316)"
                     /sex="pooled"
                     /tissue_type="heart"
                     /dev_stage="13 day embryos"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: heart; Vector: pBluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                     Oligo dT. 93 pooled NIH/Swiss 13 day embryo hearts.
                     Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
                     adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor
                     sequence: 5' CTCGAGTTTTTTTTTTTTTTTT 3'"
BASE COUNT      102 a    143 c    144 g     87 t
ORIGIN


  Query Match           7.3%;  Score 182.8;  DB 22;  Length 476;
  Best Local Similarity   64.0%;  Pred. No. 2.3e-40;
  Matches  293;  Conservative    0;  Mismatches  162;  Indels    3;  Gaps    1;

Qy   2045 ccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagttttagatgatg 2104
          |||||||||| || |||  || | | |||||||| || | ||| ||     ||| |||||
Db      2 CCTTCTACCACCAGGGCAAGGAGTTGGACTGCAGGGGAGGCCACGTGCAGCTAGCTGATG 61

Qy   2105 atacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatgatgtgtttag 2164
           ||||  || ||||||| || || || ||| | ||||| ||    ||| ||||  |||
Db     62 GCTCGGACCTGAGCTATGTGGAGGACGGCACGGCCTGTGGGCCCAACATGTTGTGCCTAG 121

Qy   2165 atcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactcgattccaagg 2224
           |||     ||||| | |      ||| | ||  | |||| ||| ||      |    ||
Db    122 ATCACCGCTGCCTGCCAGCCTCTGCCTTCAACTTCAGCACCTGCCCTGGAAGTGGAGAGC 181

Qy   2225 gtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatttgtgatttca 2284
           | |   |||| ||    ||||||||| || || || || |  |  || || ||| |
Db    182 GAAGGATCTGCTCCCATCATGGGGTTTGCAGCAACGAGGGGAAGTGTATCTGTCAGCCAG 241

Qy   2285 cctgggcagggacagattgcagtatccgggatccagt---taggaaccttcaccccccca 2341
           |||| |||| | ||| |||||||| |   | ||| |    | |  || || | |
Db    242 ACTGGACAGGCAAAGACTGCAGTATTCACAACCCACTGCCCACGTCCCCTCCCACTGGGG 301

Qy   2342 aggatgaaggacccaagggtcctagtgccaccaatctcataataggctccatcgctggtg 2401
           ||  |||  ||  ||| ||||| || |  |||||  |||| || ||||||||||| || |
Db    302 AGACTGAGAGATACAAAGGTCCCAGCGGTACCAACATCATCATTGGCTCCATCGCCGGGG 361

Qy   2402 ccatcctggtagcagctattgtccttgggggcacaggctggggatttaaaaatgtcaaga 2461
           |  |||||| ||||| || |||||  | ||||| |||||||||||||||||  ||
Db    362 CTGTCCTGGTCGCAGCCATCGTCCTGTGCGGCACGGGCTGGGGATTTAAAAACATCCGTC 421

Qy   2462 agagaaggttcgatcctactcagcaaggccccatctga 2499
           |||||| ||| || || ||||| ||  |  | |||
Db    422 GTGGAAGGTACGACCCGACCCAGCAGGGGGCAGTGTGA 459


RESULT  20
AL135156
LOCUS       AL135156      468 bp    mRNA            EST       25-FEB-2000
DEFINITION  DKFZp762L0912_r1 762 (synonym: hmel2) Homo sapiens cDNA clone
            DKFZp762L0912 5', mRNA sequence.
ACCESSION   AL135156
VERSION     AL135156.1  GI:6603343
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 468)
  AUTHORS   Ansorge,W., Wirkner,U., Mewes,W., Weil,B. and Wiemann,S.
  TITLE     EST (Ansorge,W., Wirkner,U., Mewes,H.W., Weil,B. and Wiemann,S.)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Ansorge W
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by EMBL (European Molecular Biology Laboratories,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            No s1 sequence available.
            This clone (DKFZp762L0912) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .468
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp762L0912"
                     /clone_lib="762 (synonym: hmel2)"
                     /tissue_type="melanoma (MeWo cell line)"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT      135 a    104 c    122 g    107 t
ORIGIN
```

Query Match 7.3%; Score 181.4; DB 105; Length 468;
Best Local Similarity 99.5%; Pred. No. 5.8e-40;
Matches 182; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy   2177 tacaaattcaagccctaaatatgagcagctgtccactcgattccaacggtaaagtctgtt 2236
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCCAACGGTAAAGTCTGTT 60

Qy   2237 cgggccatggggtgtgtagtaatgaagccacctgcatttgtgatttcacctgggcaggga 2296
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db     61 CGGGCCATGGGGTGTGTAGTAACGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGA 120

Qy   2297 cagattgcagtatccgggatccagttaggaaccttcacccccccaaggatgaaggaccca 2356
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 CAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAAGGATGAAGGACCCA 180

Qy   2357 agg 2359
          |||
Db    181 AGG 183
```

Search completed: August 15, 2001, 20:54:09
Job time: 4997 sec

Art No. 1

OM of: US-09-634-252A-2 to: SwissProt_39:* out_format : pfs

Date: Aug 15, 2001 10:02 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09634252/runat_14082001_142629_9857/app_query.fasta_1.2598
-DB=SwissProt_39 -QFMT=fastan -SUFFIX=n2p.rsp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-YGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-DNS=human40.cdi -LIST=90 -DOCALIGN=200 -THR_SCORE=pct:
-MAX=100 -THR_MIN=0 -ALIGN=30 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09634252_@CGN1_1_49 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-634-252A-2
Query length: 2499
Database: SwissProt_39:*
Database sequences: 93495
Database length: 34255486
Search time (sec): 52.980000

make list No. of seqs = 58%

score_list:

Sequence	Strd	Orig	zScore	EScore	Len	Documentation
SwissProt_39:AD12_HUMAN	+	969.00	1173.38	1.1e-58	909	O43184 homo sapiens (human).
SwissProt_39:DISJ_BOTJA	+	751.50	911.04	7.3e-44	571	P30431 bothrops jararaca (ja
SwissProt_39:AD08_HUMAN	+	745.00	899.68	2.2e-43	824	P78325 homo sapiens (human).
SwissProt_39:AD02_HUMAN	+	737.00	890.93	7.5e-43	735	Q99965 homo sapiens (human).
SwissProt_39:AD08_MOUSE	+	685.50	826.72	2.5e-39	826	Q05910 mus musculus (mouse).
SwissProt_39:HR1B_TRIFL	+	610.00	740.51	3.2e-34	416	P20164 trimeresurus flavovir
SwissProt_39:HRTE_CROAT	+	498.50	602.53	1.3e-26	478	P34182 crotalus atrox (weste
SwissProt_39:DISA_TRIGA	+	494.50	597.59	2.5e-26	480	P15503 trimeresurus gramineu
SwissProt_39:DISR_AGKRH	+	475.50	574.34	5.0e-25	478	P30403 agkistrodon rhodostom
SwissProt_39:HRTD_CROAT	+	368.50	444.49	9.8e-18	414	P15167 crotalus atrox (weste
SwissProt_39:ATS7_HUMAN	+	324.00	381.80	1.3e-14	997	Q9ukp4 homo sapiens (human).
SwissProt_39:ATS1_HUMAN	+	309.50	364.31	1.2e-13	967	Q9uhi8 homo sapiens (human).
SwissProt_39:ATS1_MOUSE	+	304.50	358.17	2.7e-13	968	P97857 mus musculus (mouse).
SwissProt_39:ATS6_HUMAN	+	302.00	356.20	3.9e-13	860	Q9ukp5 homo sapiens (human).
SwissProt_39:ATS1_RAT	+	298.50	350.82	6.9e-13	967	Q9wuq1 rattus norvegicus (rat)
SwissProt_39:YE94_SCHPO	+	292.00	348.74	1.7e-12	512	O13766 schizosaccharomyces p
SwissProt_39:ATS5_MOUSE	+	290.50	341.38	2.4e-12	930	Q9r001 mus musculus (mouse).
SwissProt_39:ATS8_HUMAN	+	280.00	328.91	1.3e-11	890	Q9up79 homo sapiens (human).
SwissProt_39:ATS5_HUMAN	+	265.50	310.73	1.2e-10	930	Q9una0 homo sapiens (human).
SwissProt_39:ATS8_MOUSE	+	265.00	310.37	1.3e-10	905	P57110 mus musculus (mouse).
SwissProt_39:ATS3_HUMAN	+	256.00	296.72	5.8e-10	1201	O15072 homo sapiens (human)
SwissProt_39:HR2_TRIFL	+	223.00	272.81	7.4e-08	201	P20165 trimeresurus flavoviri
SwissProt_39:FIBR_AGKCO	+	222.50	272.11	8.0e-08	203	P28891 agkistrodon contortri
SwissProt_39:ATS2_HUMAN	+	219.50	251.89	1.8e-07	1211	O95450 h adam-ts 2 precurso
SwissProt_39:HRT2_CRORU	+	212.50	259.89	3.9e-07	202	P20897 crotalus ruber ruber
SwissProt_39:ATS2_BOVIN	+	210.50	240.91	7.4e-07	1205	P79331 b adam-ts 2 precurso
SwissProt_39:TENA_CHICK	+	209.50	235.92	9.3e-07	1808	P10039 gallus gallus (chick
SwissProt_39:TENA_PIG	+	209.00	235.63	1.0e-06	1746	Q29116 sus scrofa (pig). tena
SwissProt_39:ATS4_HUMAN	+	208.50	241.83	9.4e-07	837	O75173 homo sapiens (human).
SwissProt_39:HR2A_TRIFL	+	208.00	254.38	7.8e-07	202	P14530 trimeresurus flavovir
SwissProt_39:TENX_HUMAN	+	208.00	226.09	1.4e-06	4289	P22105 homo sapiens (human)
SwissProt_39:HRL2_LACMU	+	206.50	252.63	9.9e-07	200	P22796 lachesis muta muta (b
SwissProt_39:TENA_HUMAN	+	204.50	227.97	2.1e-06	2201	P24821 homo sapiens (human)
SwissProt_39:ADAM_CROAD	+	195.50	239.01	5.6e-06	203	P34179 crotalus adamanteus (
SwissProt_39:GLP1_CAEEL	+	188.00	212.66	2.6e-05	1295	P13508 caenorhabditis elega
SwissProt_39:DLL1_HUMAN	+	186.00	215.60	3.2e-05	723	O00548 homo sapiens (human).
SwissProt_39:DLL1_MOUSE	+	185.50	215.00	3.4e-05	722	Q61483 mus musculus (mouse).
SwissProt_39:LI12_CAEEL	+	185.00	208.07	4.2e-05	1429	P14585 caenorhabditis elega
SwissProt_39:NTC4_MOUSE	+	183.00	202.67	6.1e-05	1964	P31695 mus musculus (mouse)
SwissProt_39:NTC1_MOUSE	+	183.00	200.32	6.4e-05	2531	Q01705 mus musculus (mouse)
SwissProt_39:DISI_BITAR	+	181.50	230.13	4.3e-05	83	P17497 bitis arietans (afr
SwissProt_39:CRB_DROME	+	181.50	200.04	7.9e-05	2139	P10040 drosophila melanog
SwissProt_39:TGFB_HUMAN	+	180.50	202.78	8.5e-05	1394	P22064 homo sapiens (hum
SwissProt_39:FBN2_MOUSE	+	180.50	195.97	9.7e-05	2907	Q61555 mus musculus (mou
SwissProt_39:NTC3_MOUSE	+	179.50	196.85	0.0001	2318	Q61982 mus musculus (mou
SwissProt_39:NTC1_RAT	+	179.50	196.03	0.0001	2531	Q07008 rattus norvegicus (
SwissProt_39:DLL1_RAT	+	179.00	207.14	9.5e-05	714	P97677 rattus norvegicus (r
SwissProt_39:LML1_CAEEL	+	178.50	199.30	0.0001	1557	Q18823 caenorhabditis el
SwissProt_39:NTC1_HUMAN	+	178.50	195.13	0.0001	2444	P46531 homo sapiens (hum
SwissProt_39:NOTC_DROME	+	178.00	193.58	0.0001	2703	P07207 drosophila melano
SwissProt_39:NOTC_XENLA	+	176.00	191.77	0.0002	2524	P21783 xenopus laevis (a
SwissProt_39:FBN1_MOUSE	+	175.50	189.96	0.0002	2871	Q61554 mus musculus (mou
SwissProt_39:NOTC_BRARE	+	174.50	190.25	0.0002	2437	P46530 brachydanio rerio
SwissProt_39:FBN1_BOVIN	+	172.50	186.28	0.0003	2871	P98133 bos taurus (bovin
SwissProt_39:NEL_CHICK	+	172.00	197.32	0.0003	816	Q90827 gallus gallus (chic
SwissProt_39:FBN1_HUMAN	+	171.50	185.06	0.0004	2871	P35555 homo sapiens (hum
SwissProt_39:FBP1_STRPU	+	167.50	189.34	0.0006	1064	P10079 strongylocentrotu
SwissProt_39:ITB2_HUMAN	+	166.50	191.12	0.0007	769	P05107 homo sapiens (huma
SwissProt_39:LRP_CAEEL	+	166.50	174.26	0.0010	4753	Q04833 caenorhabditis ele
SwissProt_39:LMG1_DROME	+	165.50	182.89	0.0009	1639	P15215 drosophila melano
SwissProt_39:FBN2_HUMAN	+	165.50	177.57	0.0010	2911	P35556 homo sapiens (hum
SwissProt_39:FP2_MYTGA	+	165.00	193.79	0.0008	473	Q25464 mytilus galloprovin
SwissProt_39:NEL1_HUMAN	+	165.00	188.80	0.0009	810	Q92832 homo sapiens (huma
SwissProt_39:LRP2_RAT	+	165.00	172.60	0.0012	4660	P98158 rattus norvegicus (
SwissProt_39:SERR_DROME	+	164.50	183.07	0.0011	1408	P18168 drosophila melano
SwissProt_39:LRP1_CHICK	+	164.50	172.23	0.0013	4543	P98157 gallus gallus (ch
SwissProt_39:SORL_CHICK	+	162.50	179.48	0.0015	1592	Q98930 g sortilin-relate
SwissProt_39:TGFB_RAT	+	162.50	178.81	0.0015	1712	Q00918 rattus norvegicus (
SwissProt_39:NEL2_HUMAN	+	162.00	185.06	0.0014	816	Q99435 homo sapiens (huma
SwissProt_39:ITBX_DROME	+	162.00	184.72	0.0014	846	P11584 drosophila melanog
SwissProt_39:YMV2_CAEEL	+	161.50	180.53	0.0016	1246	P34504 caenorhabditis el
SwissProt_39:TSA4_GIALA	+	161.00	185.08	0.0016	713	P21849 giardia lamblia (g
SwissProt_39:FBL1_CHICK	+	160.50	184.85	0.0017	684	O73775 gallus gallus (chi
SwissProt_39:DL_DROME	+	160.50	183.03	0.0018	833	P10041 drosophila melanogas
SwissProt_39:YLK3_CAEEL	+	160.00	174.81	0.0023	1895	P41951 caenorhabditis el
SwissProt_39:ITB2_BOVIN	+	159.50	182.54	0.0021	769	P32592 bos taurus (bovine
SwissProt_39:LRP2_HUMAN	+	158.50	164.65	0.0034	4655	P98164 homo sapiens (hum
SwissProt_39:YNX3_CAEEL	+	157.00	166.72	0.0040	3051	P34576 caenorhabditis el
SwissProt_39:DISB_TRIGA	+	156.00	200.05	0.0023	73	P17495 trimeresurus gramin
SwissProt_39:TSP4_XENLA	+	156.00	176.25	0.0037	955	Q06441 xenopus laevis (af
SwissProt_39:PCK5_MOUSE	+	155.50	169.38	0.0046	1877	Q04592 mus musculus (mou
SwissProt_39:YL_DROME	+	155.50	168.86	0.0046	1984	P98163 drosophila melanoga
SwissProt_39:YKO5_CAEEL	+	155.00	179.56	0.0040	585	P34293 caenorhabditis ele
SwissProt_39:LDVR_RAT	+	155.00	175.85	0.0043	873	P98166 rattus norvegicus (r
SwissProt_39:PCK5_BRACL	+	155.00	169.70	0.0049	1696	Q9nj15 branchiostoma cal
SwissProt_39:TS11_GIALA	+	154.00	177.12	0.0048	667	Q03185 giardia lamblia (g
SwissProt_39:LML2_CAEEL	+	154.00	161.32	0.0066	3672	Q21313 caenorhabditis el
SwissProt_39:LMA2_HUMAN	+	153.50	162.25	0.0069	3110	P24043 homo sapiens (hum
SwissProt_39:DISI_AGKHA	+	153.00	196.63	0.0037	71	P21858 agkistrodon halys b
SwissProt_39:NEL2_RAT	+	152.00	172.80	0.0068	816	Q62918 rattus norvegicus (r

seq_name: SwissProt_39:AD12_HUMAN

seq_documentation_block:
ID AD12_HUMAN STANDARD; PRT; 909 AA.
AC O43184; O60470;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ADAM 12 PRECURSOR (EC 3.4.24.-) (A DISINTEGRIN AND
DE METALLOPROTEINASE DOMAIN 12) (MELTRIN ALPHA).
GN ADAM12 OR MLTN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 12S AND 12L).
RC TISSUE=Placenta;
RX MEDLINE=98079040; PubMed=9417060;
RA Gilpin B.J., Loechel F., Mattei M.-G., Engvall E., Albrechtsen R.,
RA Wewer U.M.;
RT "A novel, secreted form of human ADAM 12 (meltrin alpha) provokes

```
RT   myogenesis in vivo.";
RL   J. Biol. Chem. 273:157-166(1998).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=98307939; PubMed=9642263;
RA   Loechel F., Gilpin B.J., Engvall E., Albrechtsen R., Wewer U.M.;
RT   "Human ADAM 12 (meltrin alpha) is an active metalloprotease.";
RL   J. Biol. Chem. 273:16993-16997(1998).
CC   -!- FUNCTION: MAY BE INVOLVED IN MYOBLAST FUSION AS WELL AS
CC       MACROPHAGE-DERIVED GIANT CELLS (MGC) AND OSTEOCLAST FORMATION FROM
CC       MONONUCLEAR PRECURSORS.
CC   -!- COFACTOR: BINDS ONE ZINC ION.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (ISOFORM 12L). A
CC       SECRETED FORM (ISOFORM 12S) IS PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 12L (SHOWN HERE) AND 12S; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING. ISOFORM 12S IS A SECRETED FORM
CC       AND DISPLAYS METALLOPROTEASE ACTIVITY.
CC   -!- TISSUE SPECIFICITY: ISOFORM 12L IS EXPRESSED IN PLACENTA AND
CC       SKELETAL, CARDIAC, AND SMOOTH MUSCLE. ISOFORM 12S SEEMS TO BE
CC       EXPRESSED ONLY IN PLACENTA OR IN EMBRYO AND FETUS. BOTH FORMS
         WERE EXPRESSED IN SOME TUMOR CELLS LINES. NOT DETECTED IN BRAIN,
         LUNG, LIVER, KIDNEY OR PANCREAS.
     -!- DOMAIN: CYS-179 WITHIN THE PROPEPTIDE ACTS AS A CYSTEINE SWITCH,
         BEING ABLE TO CHELATE THE ACTIVE SITE ZINC ION, THUS BLOCKING THE
         ACTIVITY OF THE ENZYME.
CC   -!- DOMAIN: THE CYSTEINE-RICH DOMAIN COULD BE INVOLVED IN TUMOR CELL
CC       ADHESION.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 3 SH3-BINDING DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS A DISINTEGRIN DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF023476; AAC08702.1; -.
DR   EMBL; AF023477; AAC08703.1; -.
DR   MEROPS; M12.212; -.
DR   MIM; 602714; -.
DR   HSSP; P17494; 1KST.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Transmembrane; EGF-like domain; SH3-binding; Alternative splicing.
FT   SIGNAL       1    28       POTENTIAL.
FT   PROPEP      29   207       REMOVED IN MATURE FORM.
FT   CHAIN      208   909       ADAM 12.
FT   DOMAIN     208   708       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   709   729       POTENTIAL.
FT   DOMAIN     730   909       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN     208   416       METALLOPROTEASE-LIKE.
FT   DOMAIN     417   512       DISINTEGRIN-LIKE.
FT   DOMAIN     514   649       CYS-RICH.
FT   DOMAIN     656   688       EGF-LIKE.
FT   SITE       179   179       CYSTEINE SWITCH.
FT   SITE       486   488       CELL ATTACHEMENT (POTENTIAL).
FT   ACT_SITE   351   351
FT   METAL      350   350       ZINC (CATALYTIC) (POTENTIAL).
FT   METAL      354   354       ZINC (CATALYTIC) (POTENTIAL).
FT   METAL      360   360       ZINC (CATALYTIC) (POTENTIAL).
FT   DISULFID   482   495       POTENTIAL.
FT   DISULFID   660   670       BY SIMILARITY.
FT   DISULFID   664   676       BY SIMILARITY.
FT   DISULFID   678   687       BY SIMILARITY.
FT   DOMAIN     834   840       SH3-BINDING (POTENTIAL).
FT   DOMAIN     885   891       SH3-BINDING (POTENTIAL).
FT   DOMAIN     890   896       SH3-BINDING (POTENTIAL).
FT   CARBOHYD   111   111       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   149   149       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   452   452       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   381   381       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   651   651       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   705   738       DNQGLTIGILVTILCLLAAGFVVYLKRKTLIRLL -> EAR
FT                              QEAAESNRERGQGQEPVGSQEHASTASLTLI (IN
FT                              ISOFORM 12S).
FT   VARSPLIC   739   909       MISSING (IN ISOFORM 12S).
SQ   SEQUENCE   909 AA;  99642 MW;  B74B35DFCB0A70BA CRC64;


alignment_scores:
          Quality:  969.00            Length:    809
            Ratio:  2.036               Gaps:     29
 Percent Similarity:  58.838   Percent Identity: 30.902

alignment_block:
 US-09-634-252A-2 x AD12_HUMAN   ..

 Align seg 1/1  to: AD12_HUMAN  from: 1  to: 909

    138 GCTTCTCGTCCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCC 187
        |||:::|||||   |||::::::   ||||||   |||:::   |||
      2 AlaAlaArgProLeuProValSerProAlaArgAlaLeuLeuLeuAla.L 18

    188 GCGCCTGGGGGGCTGCTGCGCCCAGCGCTCCGCAT.....TGG 225
          |||   :::      ||||||         :::          |||
     18 euAlaGlyAlaLeuLeuAlaProCysGluAlaArgGlyValSerLeuTrp 34

    226 AATGAAACTGCAGAAAAAAATTTGGGAGTCCTGGCAGATGAAGACAATAC 275
        ||||||                          |||||||||
     35 AsnGluGlyArg.....................AlaAspGlu..... 41

    276 ATTGCAACAGAATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGA 325
                                  :::     :::::::::   :::
     42 .......................ValValSerAlaSerValArgSerG 50

    326 AAGAAATCACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCG 375
          ::::::   :::|||   :::              :::   ||||||
     50 lyAspLeuTrpIleProValLys..............SerPheAspSer 61

    376 GAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACA 425
        ::::::      :::|||||:::   :::   ::::::::::::|||
     62 LysAsnHisProGluValLeuAsnIleArgLeuGlnArgGluSerLys.. 77

    426 TAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCG 475

     77 ..................................................... 77

    476 GCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCT 525
              ::::::|||:::::|||      |||:::||||||:::::|||
     78 .....GluLeuIleIleAsnLeuGluArgAsnGluGlyLeuIleAlaSer 92

    526 GATTATGTGGAGATTCACTAC...GAAAATGGGAAACCA.......... 561
        ::::::::::|||   ||||||   ::::::|||
     93 SerPheThrGluThrHisTyrLeuGlnAspGlyThrAspValSerLeuAl 109

    562 ....CAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCA 607
            :::|||:::            ||||||||||||||||||   :::|
    109 aArgAsnTyrThrValIleLeuGlyHisCysTyrTyrHisGlyHisValA 126

    608 GAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCAT 657
        |||||   :::||||||   |||:::||||||||||||:::||||||:::
    126 rgGlyTyrSerAspSerAlaValSerLeuSerThrCysSerGlyLeuArg 142
```

```
658 GGCATG......TTTGAAGATGATACCTTCGTGTATATGATAGAGCCACT 701
143 GlyLeuIleValPheGluAsnGluSer......TyrValLeuGluProMe 157

702 AGAGCTGGTTCATGATGAGAAAAGCACAGGTCGA............CCAC 739
157 t..............LysSerAlaThrAsnArgTyrLysLeuPheProA 169

740 ATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAAT 789
169 laLysLysLeuLysSerValArgGlySerCysGlySerHisHisAsnThr 185

790 CTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTG 839
186 ProAsnLeuAlaAlaLysAsnValPheProProProSerGln...ThrTr 201

840 GTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAG 889
201 pAlaArgArgHisLysArgGluThr.................LeuLysA 212

890 AAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAG 939
212 laThrLysTyrValGluLeuValIleValAlaAspAsnArgGluPheGln 228

940 AAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGT 989
229 ArgGlnGlyLysAspLeuGluLysValLysGlnArgLeuIleGluIleAl 245

990 CAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCC 1039
245 aAsnHisValAspLysPheTyrArg...ProLeuAsnIleArgIleValL 261

1040 TGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACC 1089
261 euValGlyValGluValTrpAsnAspMetAspLysCysSerValSerGln 277

1090 AACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGC..... 1134
278 AspProPheThrSerLeuHisGluPheLeuAspTrpArgLysMetLysLe 294

1135 .......ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGA 1177
294 uLeuProArgLysSerHis...AspAsnAlaGlnLeuValSerGlyValT 310

1178 CATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCT 1227
310 yrPheGlnGlyThrThrIleGlyMetAlaProIleMetSerMetCysThr 326

1228 ......CGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGC 1271
327 AlaAspGlnSerGlyGlyIleValMetAspHisSerAspAsnProLeuGl 343

1272 AGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAAT 1321
343 yAlaAlaValThrLeuAlaHisGluLeuGlyHisAsnPheGlyMetAsnH 360

1322 GGGAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGT 1371
360 isAspThrLeuAspArgGlyCysSerCysGlnMetAlaValGluLysGly 376

1372 GGCTGCATCATGGAG...GAAACAGGGGTGTCCCATTCTCGAAAATTTTC 1418
377 GlyCysIleMetAsnAlaSerThrGlyTyrProPheProMetValPheSe 393

1419 AAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAG 1468
393 rSerCysSerArgLysAspLeuGluThrSerLeuGluLysGlyMetGlyV 410

1469 CCTGCCTTTTCAACAGGCCA......ACAAAGCTATTTGAGCCCACGGAA 1512
410 alCysLeuPheAsnLeuProGluValArgGluSerPheGlyGlyGlnLys 426

1513 TGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCA 1562

427 CysGlyAsnArgPheValGluGluGlyGluGluCysAspCysGlyGluPr 443

1563 TGTGGAATGCTATGGATTATGCTGTAAG......AAATGTTCCCTCTCCA 1606
443 oGluGluCysMetAsnArgCysCysAsnAlaThrThrCysThrLeuLysP 460

1607 ACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTT 1656
460 roAspAlaValCysAlaHisGlyLeuCysCysGluAsp......CysGln 474

1657 TTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATAT 1706
475 LeuLysProAlaGlyThrAlaCysArgAspSerSerAsnSerCysAspLe 491

1707 TACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATA 1756
491 uProGluPheCysThrGlyAlaSerProHisCysProAlaAsnValTyrL 508

1757 AGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGC 1806
508 euHisAspGlyHisSerCysGlnAspValAspGlyTyrCysTyrAsnGly 524

1807 GAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGC 1856
525 IleCysGlnThrHisGluGlnGlnCysValThrLeuTrpGlyProGlyAl 541

1857 TGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTG 1906
541 aLysProAlaProGlyIleCysPheGluArgValAsnSerAlaGlyAspP 558

1907 AGAAGGGAAACTGCGGGAAG...GATGGAGACCGGTGGATTCAGTGCAGC 1953
558 roTyrGlyAsnCysGlyLysValSerLysSerSerPheAlaLysCysGlu 574

1954 AAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGC 2003
575 MetArgAspAlaLysCysGlyLysIleGlnCysGlnGlyGlyAlaSerAr 591

2004 TCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACC 2053
591 gProValIleGlyThrAsnAlaValSerIleGluThrAsnIleProLeuG 608

2054 ATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGAT 2103
608 lnGlnGlyGlyArgIleLeuCysArgGlyThrHisValTyrLeuGlyAsp 624

2104 GAT...ACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTC 2150
625 AspMetProAspProGlyLeuValLeuAlaGlyThrLysCysAlaAspGl 641

2151 TATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGA 2200
641 yLysIleCysLeuAsnArgGlnCysGlnAsnIleSerValPheGlyValH 658

2201 GCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTG 2250
658 isGluCysAlaMetGln..............CysHisGlyArgGlyVal 669

2251 TGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGA 2300
670 CysAsnAsnArgLysAsnCysHisCysGluAlaHisTrpAla........ 683

2301 TTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCC.......... 2340
684 ................................ProProPheCysAspL 689

2341 ....................AAGGATGAAGGACCC...AAGGGTCCTAGT 2367
689 ysPheGlyPheGlyGlySerThrAspSerGlyProIleArgGlnAlaAsp 705

2368 GCCACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGC 2417
```

```
     706 AsnGlnGlyLeuThrIleGly.............IleLeuValThrIl 717

    2418 TATTGTCCTTGGGGGCACAGGCTGG 2442
         :::   |||   ::::::|||:::
     717 eLeuCysLeuLeuAlaAlaGlyPhe 725

seq_name: SwissProt_39:DISJ_BOTJA

seq_documentation_block:
ID   DISJ_BOTJA     STANDARD;      PRT;   571 AA.
AC   P30431;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   PUTATIVE VENOM METALLOPROTEINASE JARARHAGIN PRECURSOR (EC 3.4.24.73)
DE   (HF2-PROTEINASE) [CONTAINS: DISINTEGRIN] (FRAGMENT).
OS   Bothrops jararaca (Jararaca).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Viperidae; Crotalinae; Bothrops.
OX   NCBI_TaxID=8724;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Venom gland;
RX   MEDLINE=93054601; PubMed=1385408;
RA   Paine M.J.I., Desmond H.P., Theakston R.D.G., Crampton J.M.;
RT   "Purification, cloning, and molecular characterization of a high
RT   molecular weight hemorrhagic metalloprotease, jararhagin, from
RT   Bothrops jararaca venom. Insights into the disintegrin gene family.";
RL   J. Biol. Chem. 267:22869-22876(1992).
CC   -!- FUNCTION: THIS PROTEIN IS A ZINC PROTEASE FROM SNAKE VENOM THAT
CC       ACTS IN HEMORRHAGE (BY SIMILARITY).
CC   -!- FUNCTION: INHIBITS FIBRINOGEN INTERACTION WITH PLATELET RECEPTORS
CC       EXPRESSED ON GLYCOPROTEIN IIB-IIIA COMPLEX. ACT BY BINDING TO THE
CC       GLYCOPROTEIN IIB-IIIA RECEPTOR ON THE PLATELET SURFACE AND INHIBIT
CC       AGGREGATION INDUCED BY ADP, THROMBIN, PLATELET-ACTIVATING FACTOR
CC       AND COLLAGEN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: CLEAVAGE OF 10-HIS-|-LEU-11, 14-ALA-|-LEU-15,
CC       16-TYR-|-LEU-17 AND 24-PHE-|-PHE-25 BONDS IN INSULIN B CHAIN.
CC   -!- COFACTOR: BINDS ONE ZINC ION.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: IN THE CENTRAL REGION; BELONGS TO THE DISINTEGRIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X68251; CAA48323.1; -.
DR   PIR; S24789; S24789.
DR   PIR; A44463; A44463.
DR   HSSP; P17494; 1KST.
DR   MEROPS; M12.138; -.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF00200; disintegrin; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hydrolase; Metalloprotease; Zinc; Blood coagulation; Platelet;
KW   Cell adhesion; Venom; Zymogen.
FT   NON_TER       1      1
FT   PROPEP       <1    151       POTENTIAL.
FT   CHAIN        152    355       JARARHAGIN.
FT   CHAIN        356    447       DISINTEGRIN.
FT   PROPEP       448    571
FT   MOD_RES      152    152       PYRROLIDONE CARBOXYLIC ACID.
FT   METAL        295    295       ZINC (CATALYTIC) (PROBABLE).
FT   ACT_SITE     296    296       BY SIMILARITY.
FT   METAL        299    299       ZINC (CATALYTIC) (PROBABLE).
FT   METAL        305    305       ZINC (CATALYTIC) (PROBABLE).
FT   DISULFID     270    350       BY SIMILARITY.
FT   DISULFID     310    317       BY SIMILARITY.
FT   DISULFID     377    395       BY SIMILARITY.
FT   DISULFID     379    390       BY SIMILARITY.
FT   DISULFID     389    412       BY SIMILARITY.
FT   DISULFID     403    409       BY SIMILARITY.
FT   DISULFID     408    434       BY SIMILARITY.
FT   DISULFID     421    441       BY SIMILARITY.
FT   CARBOHYD     333    333       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   571 AA;  63982 MW;  F5E549F8BF61177B CRC64;


alignment_scores:
            Quality:  751.50              Length:    617
              Ratio:  2.020                 Gaps:     23
 Percent Similarity:  60.292   Percent Identity: 30.146

alignment_block:
 US-09-634-252A-2 x DISJ_BOTJA   ..

 Align seg 1/1  to: DISJ_BOTJA  from: 1  to: 571

    403 AAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGC 452
        |||:::      |||   |||::::::   |||:::
      5 LysGlyAlaValGlnProLysTyrGluAspAlaMet.............. 16

    453 AAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATAC 502
         ::::::::::::::::::   |||:::      :::|||   |||
     17 .GlnTyrGluPheLysValAsnGlyGluProValValLeuHisLeuGluL 33

    503 TGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAA... 549
          |||:::||||||:::|||:::||||||   ||||||||||||:::
     33 ysAsnLysGlyLeuPheSerLysAspTyrSerGluIleHisTyrSerPro 49

    550 AATGGGAAA......CCACAGTACTCTAAGGGTGGAGAGCACTGTTACTA 593
        :::|||:::           |||           :::||||||||||
     50 AspGlyArgGluIleThrThrTyrProProValGluAspHisCysTyrTy 66

    594 CCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCT 643
        |||||||   |||::::::      ||||||   :::::::::|||:::|
     66 rHisGlyArgIleGluAsnAspAlaAspSerThrAlaSerIleSerAlaC 83

    644 GCAATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATA 693
        ||||||||||   |||   |||:::   :::       :::|||:::|||
     83 ysAsnGlyLeuLysGlyTyrPheLysLeuGlnArgGluThrTyrPheIle 99

    694 GAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCACATAT 743
        |||||||||:::|||      :::|||                  |||
    100 GluProLeuLysLeuProAspSerGlu.....AlaHisAl 111

    744 AATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCA 793
         :::                                :::|||     :
    111 aVal.............................PheLysTyrGluA 117

    794 CTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTA......CAG 837
        :::::|||:::   |||:::   |||   :::      :::      :::
    117 snValGluLysGluAspGluAlaProLysMetCysGlyValThrGlnAsn 133

    838 TGG.................TTGAAAAGAAGGAAGAGAGCAGTGAATCC 869
        |||                 |||      :::   |||   :::
    134 TrpLysSerTyrGluProIleLysLysAlaSerGlnLeuAlaPheThrAl 150

    870 ATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTA 919
         ::::::      ::::::      ||||||:::|||:::::::::|||
    150 aGluGlnGlnArgTyrAspProTyrLysTyrIleGluPhePheValValV 167
```

```
 920 ATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAAC 969
     |||:::   |||      |||:::::::::::::            :::
 167 alAspGlnGlyThrValThrLysAsnAsnGlyAspLeuAspLysIleLys 183

 970 AACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCA 1019
               :::::::::|||:::|||:::::::|||::::::
 184 AlaArgMetTyrGluLeuAlaAsnIleValAsnGluIlePheArg...Ty 199

1020 GCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGG 1069
      |||      :::|||:::||||||:::::|||   |||::::::   |
 199 rLeuTyrMetHisValAlaLeuValGlyLeuGluIleTrpSerAsnGlyA 216

1070 ATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCA 1119
     ||:::|||   :::      :::            |||:::::::|||:::
 216 spLysIleThrValLysProAspValAspTyrThrLeuAsnSerPheAla 232

1120 AAATACCGGCAG.............CGCATTAAGCAGCATGCTGATGCTGT 1157
     ::::::|||::::                |||:::|||   |||   ::
 233 GluTrpArgLysThrAspLeuLeuThrArgLysLysHis...AspAsnAl 248

1158 GCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTT 1207
     ::::|||::::::   :::   |||:::            :::   :::|
 248 aGlnLeuLeuThrAlaIleAspPheAsnGlyProThrIleGlyTyrAlaT 265

1208 ACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGT...GTGAATGAG 1254
     ||:::|||::::::|||         |||:::||||||   |||::::::
 265 yrIleGlySerMetCysHisProLysArgSerValGlyIleValGlnAsp 281

1255 TATGGT...CTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGC 1301
     |||:::   :::   ::::::||||||   ::::::::::::::::::::
 282 TyrSerProIleAsnLeuValValAlaValIleMetAlaHisGluMetGl 298

1302 TCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATGTG 1351
     ::::|||||||||||||:::   :::   :::|||            |||:
 298 yHisAsnLeuGlyIleHisHisAspThrGlySer...........CysS 311

1352 ACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCC 1401
     ::|||   :::            |||||||||      |||   :::|||
 311 erCysGlyAspTyr......ProCysIleMetGlyProThr...IleSer 324

1402 CAT......TCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGA 1445
     :::      |||:::   ||||||:::||||||   ::::::      ||
 325 AsnGluProSerLysPhePheSerAsnCysSerTyrIleGlnCysTrpAs 341

1446 CTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCA......A 1489
     ||||:::::::::   :::      |||:::::::|||:::|||      |
 341 pPheIleMetAsnHisAsnProGluCysIleIleAsnGluProLeuGlyT 358

1490 CAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGG 1539
     ||   ::::::::::|||      |||||||||      :::|||:::|||
 358 hrAspIleIleSerProProValCysGlyAsnGluLeuLeuGluValGly 374

1540 GAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTATGGATTATGCTGT.. 1587
     ||||||||||||||||||         :::|||   :::   ||||||
 375 GluGluCysAspCysGlyThrProGluAsnCysGlnAsnGluCysCysAs 391

1588 ....AAGAAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCT 1633
             |||:::|||::::::|||:::::::|||:::   |||   |
 391 pAlaAlaThrCysLysLeuLysSerGlySerGlnCysGlyHisGlyAspC 408

1634 GCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGG 1683
     ||||||:::      :::|||   |||:::      |||   |||||||||
 408 ysCysGlu......GlnCysLysPheSerLysSerGlyThrGluCysArg 422

1684 GATGCTGTGAACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGG 1733
        :::::::::|||||||||   :::|||:::|||||||||:::|||::
 423 AlaSerMetSerGluCysAspProAlaGluHisCysThrGlyGlnSerSe 439

1734 TCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAA 1783
     ::::|||||   ::::::      ::::::|||      |||   :::|
 439 rGluCysProAlaAspValPheHisLysAsnGlyGlnProCysLeuAspA 456

1784 ATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGT 1833
     ||   |||   ||||||||||||:::|||            :::||||||
 456 snTyrGlyTyrCysTyrAsnGlyAsnCysProIleMetTyrHisGlnCys 472

1834 CAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGA 1883
           ::::::|||:::   :::      ::::::      |||:::
 473 TyrAlaLeuPheGlyAlaAspValTyrGluAlaGluAspSerCysPhe.. 488

1884 AAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAG 1933
      |||   |||   :::|||:::      |||   |||   |||::::::
 489 .LysAspAsnGlnLysGlyAsnTyrTyrGlyTyrCysArgLysGluAsnG 505

1934 ACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTC 1983
       :::   |||   |||:::   :::||||||   ||||||   |||
 505 lyLysLysIleProCysAlaProGluAspValLysCysGlyArgLeuTyr 521

1984 TGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGAT 2033
     |||   :::      :::::::|||                  |||:::
 522 CysLysAsp......AsnSerPro.................GlyGlnAs 530

2034 CATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTG 2083
          |||   ::::::::::::::::
 530 nAsnProCysLysMetPheTyrSer..................... 538

2084 CCCATGTAGTTTTAGATGATGATACGGATGTGGGCTATGTAGAAGATGGA 2133
                    :::||||||         |||   |||      |||
 539 .............AsnAspAspGluHisLysGlyMetValLeuProGly 550

2134 ACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAAT 2183
     |||   |||:::   :::   :::|||   :::      |||::::::::
 551 ThrLysCysAlaAspGlyLysValCysSerAsnGlyHisCysValAspVa 567

2184 T 2184
     :
 567 l 567
```

seq_name: SwissProt_39:AD08_HUMAN

seq_documentation_block:

ID AD08_HUMAN STANDARD; PRT; 824 AA.
AC P78325;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ADAM 8 PRECURSOR (EC 3.4.24.-) (A DISINTEGRIN AND METALLOPROTEINASE
DE DOMAIN 8) (CELL SURFACE ANTIGEN MS2) (CD156 ANTIGEN).
GN ADAM8 OR MS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=97271556; PubMed=9126482;
RA Yoshiyama K., Higuchi Y., Kataoka M., Matsuura K., Yamamoto S.;
RT "CD156 (human ADAM8): expression, primary amino acid sequence, and
RT gene location.";
RL Genomics 41:56-62(1997).
CC -!- FUNCTION: POSSIBLE INVOLVEMENT IN EXTRAVASATION OF LEUKOCYTES.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: EXPRESSED ON NEUTROPHILS AND MONOCYTES.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC -!- SIMILARITY: CONTAINS A DISINTEGRIN DOMAIN.
CC -!- DATABASE: NAME=PROW; NOTE=CD guide CD156 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd156.htm".
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; D26579; BAA05626.1; -.
DR   HSSP; P17494; 1KST.
DR   MEROPS; M12.208; -.
DR   MIM; 602267; -.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF00200; disintegrin; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
     Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein;
     Transmembrane; Antigen.
     SIGNAL        1     16       POTENTIAL.
     CHAIN        17    824       ADAM 8.
     DOMAIN       17    655       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    656    676       POTENTIAL.
FT   DOMAIN      677    824       CYTOPLASMIC (POTENTIAL).
FT   METAL       334    334       ZINC (CATALYTIC) (PROBABLE).
FT   ACT_SITE    335    335       BY SIMILARITY.
FT   METAL       338    338       ZINC (CATALYTIC) (PROBABLE).
FT   METAL       344    344       ZINC (CATALYTIC) (PROBABLE).
FT   CARBOHYD     67     67       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD     91     91       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    436    436       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    612    612       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   824 AA;  88673 MW;  5DF8E05F30DF479E CRC64;


alignment_scores:
            Quality:  745.00              Length:    693
              Ratio:  1.961                Gaps:     29
 Percent Similarity:  54.834   Percent Identity: 30.447

alignment_block:
 US-09-634-252A-2 x AD08_HUMAN   ..

 Align seg 1/1  to: AD08_HUMAN  from: 1  to: 824

     322 CAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACATCAACCAAGA 371
         |||   |||::::::||||||   ||||||
      28 GlnTyrGluValValLeuProArgArgLeu..................... 37

     372 CTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAA 421
             :::|||                  :::::::|||:::
      38 ....ProGlyPro.................ArgValArgArgAlaLeuP 47

     422 AACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCC 471
             :::   ::::::|||      :::::::|||:::   :::   |||
      47 roSerHisLeuGlyLeuHisProGluArgValSerTyrValLeuGlyAla 63

     472 TTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTC 521
           |||   :::|||   |||   |||      |||:::   ||||||::
      64 ThrGlyHisAsnPheThrLeuHisLeuArgLysAsnArgAspLeuLeuGl 80

     522 TTCTGATTATGTGGAGATTCAC...TACGAAAATGGGAAA......CCAC 562
         :|||   |||:::|||   :::         ||||||:::         :
      80 ySerGlyTyrThrGluThrTyrThrAlaAlaAsnGlySerGluValThrG 97

     563 AGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGC 612
         ::      :::|||   :::||||||   |||:::|||   :::::::|||
      97 luGlnProArgGlyGlnAspHisCysLeuTyrGlnGlyHisValGluGly 113

     613 GTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCAT 662
         ||||||   ::::::||||||||||||   ||||||:::|||::
     114 TyrProAspSerAlaAlaSerLeuSerThrCysAlaGlyLeuArgGlyPh 130

     663 GTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTC 712
         :|||:::      :::   :::::::::||||||||||
     130 ePheGlnValGlySerAspLeuHisLeuIleGluProLeu..... 143

     713 ATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCA 762
           ||||||   :::   |||      |||   :::
     144 ..AspGluGlyGlyGluGlyGlyArgHisAlaValTyr..... 155

     763 GGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAGGT..... 807
                     |||   :::::::|||   :::      |||
     156 ..............GlnAlaGluHisLeuLeuGlnThrAlaGlyThrCy 167

     808 .............GACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGA 844
                     |||:::         |||:::::|||      |||
     167 sGlyValSerAspAspSer.........LeuGlySerLeu......LeuG 179

     845 AAAGAAGGAAGAGAGCAGTG.......................AATCCA 870
           |||       ||||||                         |||
     179 lyProArgThrAlaAlaValPheArgProArgProGlyAspSerLeuPro 195

     871 TCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAA 920
         ||||||            |||   :::|||:::||||||   :::|||
     196 SerArg............GluThrArgTyrValGluLeuTyrValValVa 208

     921 TGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAACA 970
          |||:::      ::::::         :::|||:::|||   :::::::
     208 lAspAsnAlaGluPheGlnMet...LeuGlySerGluAlaAlaValArgH 224

     971 ACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAG 1020
         ::   :::   :::|||||||||   ||||||::::::|||   ::::::
     224 isArgValLeuGluValValAsnHisValAspLysLeuTyr...GlnLys 239

    1021 CTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGA 1070
         ||||||   ||||||||||||||||:::::::|||   |||:::::::::||
     240 LeuAsnPheArgValValLeuValGlyLeuGluIleTrpAsnSerGlnAs 256

    1071 TCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAA 1120
         |:::::   :::::   :::|||         |||:::::::::
     256 pArgPheHisValSerProAspProSerValThrLeuGluAsnLeuLeuT 273

    1121 AATAC......CGGCAGCGCATTAAGCAGCATGCT...GATGCTGTGCAC 1161
           :::      ||||||||   :::::::|||      |||   |||:::
     273 hrTrpGlnAlaArgGlnArgThrArgArgHisLeuHisAspAsnValGln 289

    1162 CTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTT 1211
         ||||||:::   |||   |||                 :::::::::
     290 LeuIleThrGlyValAspPheThrGlyThrThrValGlyPheAlaArgVa 306

    1212 TGGAGGTGTCTGTTCTCGCACAAGAGGA...GTTGGTGTGAATGAGTATG 1258
          :::::::::||||||::::::   |||   |||:::   ::::::
     306 lSerAlaMetCysSerHisSerSerGlyAlaValAsnGlnAspHisSerL 323

    1259 GTCTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAAC 1308
              |||::::::||||||   :::::::::::::::::::::::::|||
     323 ysAsnProValGlyValAlaCysThrMetAlaHisGluMetGlyHisAsn 339

    1309 CTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATGTGACTGCAC 1358
         ||||||::::::   :::   :::   :::           |||   |||
     340 LeuGlyMetAspHisAspGluAsnValGln......GlyCysArgCysGl 354

    1359 AGAATCCTGG......GGTGGCTGCATCATGGAG...GAAACAGGGGTGT 1399
          |||   :::      |||   |||||||||      :::   |||   |
     354 nGluArgPheGluAlaGlyArgCysIleMetAlaGlySerIleGlySerS 371

    1400 CCCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTT 1449
         ||      |||   ||||||   ||||||             :::::|||
     371 erPheProArgMetPheSerAspCysSerGlnAlaTyrLeuGluSerPhe 387

    1450 TTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCA......ACAAA 1493
         |||:::|||      ::::::||||||   |||   |||      :::
```

```
 388 LeuGluArgProGlnSerValCysLeuAlaAsnAlaProAspLeuSerHi 404

1494 GCTATTTGAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGG 1543
      |||               |||||||||   :::||||||   ||||||:
 404 sLeuValGlyGlyProValCysGlyAsnLeuPheValGluArgGlyGluG 421

1544 AGTGTGATTGTGGTTTTCATGTGGAATGCTATGGATTATGCTGTAAGAAA 1593
     ::|||||||||||||         :::|||   :::   ||||||::::::
 421 lnCysAspCysGlyProProGluAspCysArgAsnArgCysCysAsnSer 437

1594 ......TGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTG 1637
           |||:::|||::::::||||||:::|||:::   |||   |||||
 438 ThrThrCysGlnLeuAlaGluGlyAlaGlnCysAlaHisGlyThrCysCy 454

 638 TAACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATG 1687
     |:::      :::|||      :::|||   |||      ||||||
 454 sGln......GluCysLysValLysProAlaGlyGluLeuCysArgProL 469

1688 CTGTGAACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAG 1737
       :::   ||||||:::   |||:::|||   |||          :::
 469 ysLysAspMetCysAspLeuGluGluPheCysAspGlyArgHisProGlu 485

1738 TGCCCACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCA 1787
     ||||||   :::      :::::::::|||      |||      :::
 486 CysProGluAspAlaPheGlnGluAsnGlyThrProCys......SerGl 500

1788 GGGCCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGT 1837
      |||   ||||||||||||   |||   |||      :::|||||||
 500 yGlyTyrCysTyrAsnGlyAlaCysProThrLeuAlaGlnGlnCysGlnA 517

1838 ACATCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAG 1887
       :::|||||      :::   ::::::::::::   |||:::::
 517 laPheTrpGlyProGlyGlyGlnAlaAlaGluGluSerCysPheSerTyr 533

1888 CTGAATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCG 1937
                |||   :::   :::::         |||
 534 AspIleLeuProGlyCysLysAlaSerArgTyrArgAlaAsp..... 547

1938 GTGGATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTA 1987
                              :::||||||   |||   |||
 548 ..........................MetCysGlyValLeuGlnCysL 555

1988 CCAATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATT 2037
       :::                    ||||||
 555 ysGly.....................GlyGln..... 558

2038 CCAACTTCCTTCTACCATCAAGGCCGG.........GTGATTGACTGCAG 2078
                           ||||||         ::::::|||
 559 ............GlnProLeuGlyArgAlaIleCysIleValAspValCy 571

2079 TGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTATGTAGAAG 2128
       |||   ::::::   :::|||   |||              |||   :
 571 sHisAla...LeuThrThrGluAspGlyThrAlaTyrGluProValProG 587

2129 ATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTA 2178
     ::||||||   ||||||||||:::   :::|||         :::|||
 587 luGlyThrArgCysGlyProGluLysValCysTrpLysGlyArgCysGln 603

2179 CAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAA 2228
     ::::::::::::   :::   |||:::|||      :::
 604 AspLeuHisValTyrArgSerSerAsnCysSerAlaGln..... 616

2229 AGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTG 2278
       |||   :::||||||||||||:::::::::      |||   |||
 617 ....CysHisAsnHisGlyValCysAsnHisLysGlnGluCysHisCysH 632

2279 ATTTCACCTGGGCAGGGACAGATTGCAGT 2307
             ||||||         |||:::
 632 isAlaGlyTrpAlaProProHisCysAla 641

seq_name: SwissProt_39:AD02_HUMAN

seq_documentation_block:

ID   AD02_HUMAN     STANDARD;      PRT;   735 AA.
AC   Q99965; P78326;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   ADAM 2 PRECURSOR (A DISINTEGRIN AND METALLOPROTEINASE DOMAIN 2)
DE   (FERTILIN BETA SUBUNIT) (PH-30) (PH30).
GN   ADAM2 OR FTNB.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=97193554; PubMed=9041139;
RA   Vidaeus C.M., von Kap-Herr C., Golden W.L., Eddy R.L., Shows T.B.,
RA   Herr J.C.;
RT   "Human fertilin beta: identification, characterization, and
RT   chromosomal mapping of an ADAM gene family member.";
RL   Mol. Reprod. Dev. 46:363-369(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=96295488; PubMed=8702389;
RA   Gupta S.K., Alves K., Palladino L.O., Mark G.E., Hollis G.F.;
RT   "Molecular cloning of the human fertilin beta subunit.";
RL   Biochem. Biophys. Res. Commun. 224:318-326(1996).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=97224507; PubMed=9070941;
RA   Burkin H.R., Burkin D.J., Davey P.M., Griffin D.K., Affara N.A.;
RT   "Mapping, sequence, and expression analysis of the human fertilin beta
RT   gene (FTNB).";
RL   Genomics 40:190-192(1997).
CC   -!- FUNCTION: SPERM SURFACE MEMBRANE PROTEIN THAT MAY BE INVOLVED IN
CC       SPERM-EGG PLASMA MEMBRANE ADHESION AND FUSION DURING
CC       FERTILIZATION. COULD HAVE A DIRECT ROLE IN SPERM-ZONA BINDING OR
CC       MIGRATION OF SPERM FROM THE UTERUS INTO THE OVIDUCT. INTERACTIONS
CC       WITH EGG MEMBRANE COULD BE MEDIATED VIA BINDING BETWEEN ITS
CC       DISINTEGRIN-LIKE DOMAIN TO ONE OR MORE INTEGRINS RECEPTORS ON THE
CC       EGG. THIS IS A NON CATALYTIC METALLOPROTEASE-LIKE PROTEIN.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN SPERMATOGENIC CELLS
CC       IN THE SEMINIFEROUS CELLS. NOT DETECTED IN FETAL TISSUES.
CC   -!- DOMAIN: A CONSERVED TRIPEPTIDE MOTIF (FEE) WITHIN DISINTEGRIN-LIKE
CC       DOMAIN COULD BE INVOLVED IN THE BINDING TO EGG INTEGRIN RECEPTOR
CC       AND THUS COULD MEDIATE SPERM/EGG BINDING.
CC   -!- PTM: THE SIGNAL AND THE METALLOPROTEASE DOMAIN ARE CLEAVED DURING
CC       THE EPIDIDYMAL MATURATION OF THE SPERMATOZOA.
CC   -!- MISCELLANEOUS: IN MAMMALS, EXISTS AS A HETERODIMER COMPOSED OF AN
CC       ALPHA AND BETA SUBUNITS. IN HUMAN, FERTILIN ALPHA IS A PSEUDOGENE.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS A DISINTEGRIN DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U52370; AAC51110.1; -.
DR   EMBL; U38805; AAD04206.1; -.
DR   EMBL; X99374; CAA67753.1; -.
DR   HSSP; P17494; 1KST.
DR   MEROPS; M12.973; -.
DR   MIM; 601533; -.
DR   Pfam; PF00200; disintegrin; 1.
```

```
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep;. 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00022; EGF_1; FALSE_NEG.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW   Signal; Glycoprotein; Transmembrane; EGF-like domain.
FT   SIGNAL        1     16       POTENTIAL.
FT   PROPEP       17    174       REMOVED IN MATURE FORM.
FT   CHAIN       175    735       ADAM 2.
FT   DOMAIN      175    686       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    687    707       POTENTIAL.
FT   DOMAIN      708    735       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      175    382       METALLOPROTEASE-LIKE.
FT   DOMAIN      383    475       DISINTEGRIN-LIKE.
FT   DOMAIN      477    606       CYS-RICH.
FT   DOMAIN      612    686       EGF-LIKE.
FT   DISULFID    445    458       POTENTIAL.
     DISULFID    616    627       BY SIMILARITY.
     DISULFID    621    633       BY SIMILARITY.
     DISULFID    635    644       BY SIMILARITY.
     CARBOHYD    122    122       N-LINKED (GLCNAC. . .) (POTENTIAL).
     CARBOHYD    220    220       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    353    353       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    459    459       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    566    566       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT      3      3       MISSING (IN REF. 2).
FT   CONFLICT     21     21       D -> H (IN REF. 3).
FT   CONFLICT     99     99       E -> D (IN REF. 3).
FT   CONFLICT    106    106       V -> G (IN REF. 3).
FT   CONFLICT    170    170       V -> A (IN REF. 2).
FT   CONFLICT    288    288       H -> D (IN REF. 2 AND 3).
FT   CONFLICT    321    321       T -> I (IN REF. 2 AND 3).
FT   CONFLICT    388    388       G -> S (IN REF. 3).
FT   CONFLICT    396    398       EEC -> DEF (IN REF. 3).
FT   CONFLICT    501    501       G -> S (IN REF. 3).
FT   CONFLICT    529    529       D -> Y (IN REF. 3).
FT   CONFLICT    579    579       S -> G (IN REF. 3).
FT   CONFLICT    588    588       W -> L (IN REF. 3).
FT   CONFLICT    603    603       N -> D (IN REF. 3).
FT   CONFLICT    629    630       NK -> KQ (IN REF. 3).
FT   CONFLICT    638    638       S -> F (IN REF. 3).
SQ   SEQUENCE   735 AA;  82466 MW;  BBF27B8BEC7F92D3 CRC64;


alignment_scores:
            Quality:  737.00              Length:    695
              Ratio:  1.856                Gaps:     23
 Percent Similarity:  57.122   Percent Identity: 27.482

[redacted]ment_block:
[redacted]-634-252A-2 x AD02_HUMAN   ..

[redacted]gn seg 1/1  to: AD02_HUMAN  from: 1  to: 735

     313 AATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACAT 362
         :::::::::      :::||||||:::|||:::::::::          ||
      21 AspSerLeuProValGlnIleThrValProGluLysIleArgSerIleIl 37

     363 CAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGAC 412
         |:::::      ||||||
      37 eLysGluGlyIleGluSer............................... 43

     413 ACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAG 462
                                         |||||||||::::::
      44 .................................GlnAlaSerTyrLys 48

     463 ATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGG 512
         |||         |||:::   :::   ::::::|||:::   :::|||
      49 IleValIleGluGlyLysProTyrThrValAsnLeuMetGlnLysAsnPh 65

     513 TTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAAATGGG......A 556
          |||       :::       |||       :::       :::|||      |
      65 eLeuProHisAsnPheArgValTyrSerTyrSerGlyThrGlyIleMetL 82

     557 AACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATC 606
         |||||       :::          :::   |||:::|||:::|||   |||
      82 ysProLeuAspGlnAspPheGlnAsnPheCysHisTyrGlnGlyTyrIle 98

     607 AGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCA 656
         :::|||          |||   |||   :::||||||||||:::||||||::
      99 GluGlyTyrProLysSerValValMetValSerThrCysThrGlyLeuAr 115

     657 TGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGC 706
         :|||:::::::::   ::::::       |||   ||||||||||||
     115 gGlyValLeuGlnPheGluAsnValSerTyrGlyIleGluProLeu.... 130

     707 TGGTTCATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAACC 756
                  |||:::|||:::|||      |||:::|||   ::::::
     131 ..........GluSerSerValGlyPheGluHisValIleTyrGlnVal 143

     757 TTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAGG 806
               ::::::::::|||         :::|||     .|||:::
     144 .........LysHisLysLysAlaAspValSerLeuTyrAsnGluLysAs 157

     807 TGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAAGA 856
                                  |||   :::   |||:::   :::   :
     157 p..................IleGluSerArgAspLeuSerPheLysLeuG 168

     857 GAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATTTGGAA 906
         :::::|||:::|||::::::          :::      |||||||:::|||
     168 lnSerValGluProGlnGln.........AspPheAlaLysTyrIleGlu 181

     907 CTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCA 956
         :::   ::::::   :::   :::   |||:::          |||:::
     182 MetHisValIleValGluLysGlnLeuTyrAsnHisMetGlySerAspTh 198

     957 TGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTA 1006
          :::   :::   :::   :::   :::::::::::::|||:::::::::|
     198 rThrValValAlaGlnLysValPheGlnLeuIleGlyLeuThrAsnAlaI 215

    1007 TTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACC 1056
         ||:::  .   :::::::|||      :::::::|||   :::::::|||
     215 lePhe...ValSerPheAsnIleThrIleIleLeuSerSerLeuGluLeu 230

    1057 TGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAGATGCT 1106
         |||   ::::::::::::|||      1||   :::      ::::::||
     231 TrpIleAspGluAsnLysIleAlaThrThrGlyGluAlaAsnGluLeuLe 247

    1107 CCATGAGTTCTCAAAATACCGGCAGCGC.........ATTAAGCAGCATG 1147
         ||||   |||   :::::::::                ::::::   |||
     247 uHisThrPheLeuArgTrpLysThrSerTyrLeuValLeuArgProHisA 264

    1148 CTGATGCTGTGCACCTCATCTCGCGG.....................GTG 1176
              |||      |||:::   |||                      :::
     264 spValAlaPheLeuLeuValTyrArgGluLysSerAsnTyrValGlyAla 280

    1177 ACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTC 1226
         ||||||:::   |||            :::|||   |||||||||
     281 ThrPheGlnGlyLysMetCysHisAlaAsnTyrAlaGlyGlyValValLe 297

    1227 TCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGG 1276
          :::   |||   ::::::::::::                      ::::::|
     297 uHisProArgThrIleSerLeuGlu.................SerLeuA 308

    1277 CACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAA 1326
         ||   :::|||:::|||   |||:::   ::::::|||      ::::::
     308 laValIleLeuAlaGlnLeuLeuSerLeuSerMetGlyThrThrTyrAsp 324

    1327 CCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTG 1376
           :::         ||||||:::|||      |||   :::   ||
     325 AspIleAsn.........LysCysGlnCys......SerGlyAlaValCy 336

    1377 CATCATGGAGGAAACAGGGGTGTCCCATTCT......CGAAAATTTTCAA 1420
```

```
     |||||||:::      ::::::      |||      :::   ||||||:
 336 sIleMetAsnProGluAlaIleHisPheSerGlyValLysIlePheSerA 353

1421 AGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCC 1470
     ::||||||:::   ::::::      |||:::::::::      :::
 353 snCysSerPheGluAspPheAlaHisPheIleSerLysGlnLysSerGln 369

1471 TGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCC..............  1506
     ||||||   |||:::|||         :::::::|||
 370 CysLeuHisAsnGlnProArg......LeuAspProPhePheLysGlnGl 384

1507 .ACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTG 1555
      :::   |||||||||:::   :::||||||||||||||||||||||||
 384 nAlaValCysGlyAsnAlaLysLeuGluAlaGlyGluGluCysAspCysG 401

 556 GTTTTCATGTGGAATGCTATGGATTA.........TGCTGT......AAG 1590
     ||   :::   :::|||       :::          ||||||
 401 lyThrGluGlnAspCysAlaLeuIleGlyGluThrCysCysAspIleAla 417

1591 AAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAA 1640
        |||   ::::::   |||::::::|||::::::||||||||||||::
 418 ThrCysArgPheLysAlaGlySerAsnCysAlaGluGlyProCysCysGl 434

1641 CAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTG 1690
     :|||      |||||||||:::   :::         ||||||   :::
 434 uAsn......CysLeuPheMetSerLysGluArgMetCysArgProSerP 449

1691 TGAACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGC 1740
       :::||||||||||:::   |||||||||:::|||:::|||::::::  ||
 449 heGluGluCysAspLeuProGluTyrCysAsnGlySerSerAlaSerCys 465

1741 CCACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGG 1790
     |||   |||   :::   |||   |||:::   |||:::   |||||
 466 ProGluAsnHisTyrValGlnThrGlyHisProCysGlyLeuAsnGlnTr 482

1791 CCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACA 1840
        |||   :::|||   |||   :::   |||:::||||||
 482 pIleCysIleAspGlyValCysMetSerGlyAspLysGlnCysThrAspT 499

1841 TCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTG 1890
       :::|||   ::::::      :::   :::   |||||||:::   |||
 499 hrPheGlyLysGluValGluPheGlyProSerGluCysTyrSerHisLeu 515

1891 AATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTG 1940
     |||::::::          :::||||||||||||   :::   :::   ::
 516 AsnSerLysThrAspValSerGlyAsnCysGlyIleSerAspSerGlyTy 532

1941 GATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCA 1990
     :   ||||||:::      ::::::   ||||||   |||:::|||
 532 rThrGlnCysGluAlaAspAsnLeuGlnCysGlyLysLeuIleCysLysT 549

1991 AT............CTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGT 2028
                |||   :::   ||||||      :::   :::
 549 yrValGlyLysPheLeuLeuGlnIleProArgAlaThrIleIleTyrAla 565

2029 GAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAG 2078
     :::|||                |||         :::::::::   ::
 566 AsnIleSerGly............HisLeuCysIleAlaValGluPheAl 578

2079 TGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTATGTAGAAG 2128
     ::::   |||:::            |||::::::      :::::::::|
 578 aSerAspHisAla............AspSerGlnLysMetTrpIleLysA 591

2129 ATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTA 2178
     ||||||||   ||||||   :::   :::|||   :::::::::|||:::
 591 spGlyThrSerCysGlySerAsnLysValCysArgAsnGlnArgCysVal 607

2179 CAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAA 2228
     :::   :::   |||:::      :::|||      |||:::
 608 SerSerSerTyrLeuGlyTyr...AspCysThrThrAspLys........ 620

2229 AGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTG 2278
        |||:::   :::||||||||||:::|||:::      |||   |||:
```

```
 621 ....CysAsnAspArgGlyValCysAsnAsnLysLysHisCysHisCysS 636

2279 ATTTCACCTGGGCAGGGACAGATTGCAGTATCCGG 2313
     ::   ::::::         |||||||||::::::
 636 erAlaSerTyrLeuProProAspCysSerValGln 647
```

```
seq_name: SwissProt_39:AD08_MOUSE

seq_documentation_block:
ID   AD08_MOUSE      STANDARD;      PRT;   826 AA.
AC   Q05910;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   ADAM 8 PRECURSOR (EC 3.4.24.-) (A DISINTEGRIN AND METALLOPROTEINASE
DE   DOMAIN 8) (CELL SURFACE ANTIGEN MS2) (MACROPHAGE CYSTEINE-RICH
DE   GLYCOPROTEIN) (CD156 ANTIGEN).
GN   ADAM8 OR MS2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR;
RA   Yamamoto S., Yoshiyama K., Setoguchi M., Matsuura K., Higuchi Y.,
RA   Akizuki S.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=Liver;
RX   MEDLINE=97364747; PubMed=9218457;
RA   Kataoka M., Yoshiyama K., Matsuura K., Hijiya N., Higuchi Y.,
RA   Yamamoto S.;
RT   "Structure of the murine CD156 gene, characterization of its
RT   promoter, and chromosomal location.";
RL   J. Biol. Chem. 272:18209-18215(1997).
RN   [3]
RP   PRELIMINARY SEQUENCE FROM N.A.
RC   STRAIN=ICR;
RX   MEDLINE=91197896; PubMed=1982220;
RA   Yoshida S., Setoguchi M., Higuchi Y., Akizuki S., Yamamoto S.;
RT   "Molecular cloning of cDNA encoding MS2 antigen, a novel cell surface
RT   antigen strongly expressed in murine monocytic lineage.";
RL   Int. Immunol. 2:585-591(1990).
CC   -!- FUNCTION: POSSIBLE INVOLVEMENT IN EXTRAVASATION OF LEUKOCYTES.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC   -!- TISSUE SPECIFICITY: MACROPHAGES.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS A DISINTEGRIN DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D10911; BAA21771.1; -.
DR   EMBL; X13335; CAA31712.1; -.
DR   HSSP; P18619; 1FVL.
DR   MEROPS; M12.208; -.
DR   MGD; MGI:107825; Adam8.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF00200; disintegrin; 1.
```

```
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Transmembrane; Glycoprotein; Antigen; Zinc; Hydrolase;
KW   Metalloprotease; Signal.
FT   SIGNAL        1    16       POTENTIAL.
FT   CHAIN        17   826       ADAM 8.
FT   DOMAIN       17   658       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    659   683       POTENTIAL.
FT   DOMAIN      684   826       CYTOPLASMIC (POTENTIAL).
FT   METAL       329   329       ZINC (CATALYTIC) (PROBABLE).
FT   ACT_SITE    330   330       BY SIMILARITY.
FT   METAL       333   333       ZINC (CATALYTIC) (PROBABLE).
FT   METAL       339   339       ZINC (CATALYTIC) (PROBABLE).
FT   CARBOHYD     89    89       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    260   260       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    431   431       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    614   614       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   826 AA;  90046 MW;  3142CC81DBBADFB9 CRC64;

alignment_scores:
            Quality:  685.50              Length:    733
              Ratio:  1.731                 Gaps:     29
 Percent Similarity:  54.025   Percent Identity: 29.332

alignment_block:
 US-09-634-252A-2 x AD08_MOUSE   ..

 Align seg 1/1  to: AD08_MOUSE  from: 1  to: 826

     322 CAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACATCAACCAAGA 371
          |||   |||::::::   |||   ||||||
      26 GlnTyrGluValValTrpProArgArgLeu..................... 35

     372 CTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAA 421
                                 :::::::::|||:::
      36 .....................AlaAlaSerArgSerArgArgAlaLeuP 45

     422 AACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCC 471
              :::   :::   :::       :::   |||:::   :::   :::
      45 roSerHisTrpGlyGlnTyrProGluSerLeuSerTyrAlaLeuGlyThr 61

     472 TTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTC 521
            |||       |||   |||   |||       |||:::   ||||||::
      62 SerGlyHisValPheThrLeuHisLeuArgLysAsnArgAspLeuLeuGl 78

     522 TTCTGATTATGTGGAGATTCAC...TACGAAAATGGGAAA......CCAC 562
          :|||:::|||:::|||   :::         ||||||:::        :
      78 ySerSerTyrThrGluThrTyrSerAlaAlaAsnGlySerGluValThrG 95

      63 AGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGC 612
         ::      :::      :::||||||   |||:::|||   ::::::|||
      95 luGlnLeuGlnGluGlnAspHisCysLeuTyrGlnGlyHisValGluGly 111

     613 GTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCAT 662
            :::   |||   :::::::::|||||||||   ||||||:::|||::
     112 TyrGluGlySerAlaAlaSerIleSerThrCysAlaGlyLeuArgGlyPh 128

     663 GTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTC 712
          :|||:::      :::   |||:::::|||||||||||||:::
     128 ePheArgValGlySerThrValHisLeuIleGluProLeuAspAla.... 143

     713 ATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCA 762
           ||||||:::
     144 ..AspGluGlu..................................... 146

     763 GGACAGTATTCTAAG...CAAATGAAGAATCTCACTATGGAAAGAGGT.. 807
         ||||||::::::      |||   |||:::|||   ::::::   |||
     147 GlyGlnHisAlaMetTyrGlnAlaLysHisLeuGlnGlnLysAlaGlyTh 163

     808 .........................GACCAGTGGCCCTTTCTCTCTG 829
                                   |||      |||         |
     163 rCysGlyValLysAspThrAsnLeuAsnAspLeuGlyProArgAlaLeuG 180

     830 AATTA.................CAGTGGTTGAAAAGAAGGAAGAGAGCA 861
         ||:::                    :::||||||
     180 luIleTyrArgAlaGlnProArgAsnTrpLeu..................... 190

     862 GTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTTAT 911
                          |||   :::|||   :::|||:::||||||
     191 .................IleProArgGluThrArgTyrValGluLeuTy 201

     912 GATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCAC 961
          :::|||   |||   :::   ::::::|||      |||   :::|||
     201 rValValAlaAspSerGlnGluPheGlnLysLeuGlySerArgGluAla. 217

     962 ATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTAC 1011
           :::::::::   :::   :::|||||||||   ||||||::::::|||
     218 ..ValArgGlnArgValLeuGluValValAsnHisValAspLysLeuTyr 233

    1012 AAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGAC 1061
         :::|||   |||:::   ||||||||||||||||:::::|||   |||
     234 GlnGlu...LeuSerPheArgValValLeuValGlyLeuGluIleTrp.. 248

    1062 TGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATG 1111
          :::||||||::::::   |||:::                    |||::::
     249 .AsnLysAspLysPheTyrIleSerArgTyrAlaAsnValThrLeuGluA 265

    1112 AGTTCTCAAAATACCGGCAGCGC......ATTAAGCAGCATGCT...GAT 1152
         ::|||   :::::::|||::::::          ||||||      |||
     265 snPheLeuSerTrpArgGluGlnAsnLeuGlnGlyGlnHisProHisAsp 281

    1153 GCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCT 1202
            |||:::||||||:::   |||   |||      :::      :::||
     282 AsnValGlnLeuIleThrGlyValAspPheIleGlySerThrValGlyLe 298

    1203 GAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGA...GTTGGTGTGA 1249
         |:::      :::::::::|||||||||      |||   |||:::   :
     298 uAlaLysValSerAlaLeuCysSerArgHisSerGlyAlaValAsnGlnA 315

    1250 ATGAGTATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTG 1299
         :::::            :::::::||||||:::::::::::::::::::|||
     315 spHisSerLysAsnSerIleGlyValAlaSerThrMetAlaHisGluLeu 331

    1300 GCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATG 1349
         ::::::|||||||||::::::   :::   :::          |||   ||
     332 GlyHisAsnLeuGlyMetSerHisAspGluAspIle......ProGlyCy 346

    1350 TGACTGCACAGAA......TCCTGGGGTGGCTGCATCATGGAGGAAACA. 1392
         |   |||   |||       :::   ||||||||||||||   |||:::
     346 sTyrCysProGluProArgGluGlyGlyGlyCysIleMetThrGluSerI 363

    1393 ..GGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTAT 1440
           |||   :::     |||   ||||||:::||||||   ::::::
     363 leGlySerLysPheProArgIlePheSerArgCysSerLysIleAspLeu 379

    1441 AGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCA.. 1488
         ::::::|||:::   :::          :::||||||   |||   |||
     380 GluSerPheValThrLysProGlnThrGlyCysLeuThrAsnValProAs 396

    1489 ....ACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATACGTGGAAG 1534
             :::::::::                |||||||||   :::||||||
     396 pValAsnArgPheValGlyGlyProValCysGlyAsnLeuPheValGluH 413

    1535 CTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTATGGATTATGC 1584
           ||||||:::|||||||||||||          :::|||   :::   |||
     413 isGlyGluGlnCysAspCysGlyThrProGlnAspCysGlnAsnProCys 429

    1585 TGTAAG......AAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGG 1628
         |||:::          |||:::|||   :::|||||||:::|||:::::||
     430 CysAsnAlaThrThrCysGlnLeuValLysGlyAlaGluCysAlaSerGl 446
```

```
1629 GCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAAT 1678
446 yThrCysCysHis......GluCysLysValLysProAlaGlyGluValC 461

1679 GCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTACTGGAGAC 1728
461 ysArgLeuSerLysAspLysCysAspLeuGluGluPheCysAspGlyArg 477

1729 TCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATGCATGCAA 1778
478 LysProThrCysProGluAspAlaPheGlnGlnAsnGlyThrProCys.. 493

1779 TCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAACC 1828
494 ....ProGlyGlyTyrCysPheAspGlySerCysProThrLeuAlaGlnG 509

829 AGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGC 1878
509 lnCysArgAspLeuTrpGlyProGlyAlaArgValAlaAlaAspSerCys 525

1879 TATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGA 1928
526 Tyr.............ThrPheSerIleProProGlyCysAsnGlyArg.. 537

1929 TGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCT 1978
538 ................MetTyrSerGlyArgIleAsnArgCysGlyAlaL 549

1979 TACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGT 2028
549 euTyrCysGluGly.....................GlyGlnLysProLeu 558

2029 GAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAG 2078
559 GluArgSerPheCysThrPheSerSerAsnHisGlyValCysHisAlaLe 575

2079 TGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTATGTAGAAG 2128
575 uGlyThrGlySerAsnIleAsp.........ThrPheGluLeuValLeuG 589

2129 ATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTA 2178
589 lnGlyThrLysCysGluGluGlyLysValCysMetAspGlySerCysGln 605

2179 CAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAA 2228
606 AspLeuArgValTyrArgSerGluAsnCys..............SerAl 617

2229 AGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTG 2278
617 aLysCysAsnAsnHisGlyValCysAsnHisLysArgGluCysHisCysH 634

2279 ATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGTTAGGAAC 2328
634 isLysGlyTrpAlaProProAsnCys..............ValGlnArg 645

2329 CTTCACCCCCCCAAGGATGAAGGACCCAAGGGTCCTAGTGCCACCAATCT 2378
646 LeuAlaAspValSerAspGluGlnAlaAlaSerThrSerLeuProValSe 662

2379 CATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCTATTGTCCTT 2427
662 rValVal......ValValLeuValIleLeuValAlaAlaMetValIle 676
```

```
seq_name: SwissProt_39:HR1B_TRIFL

seq_documentation_block:
ID   HR1B_TRIFL     STANDARD;      PRT;   416 AA.
AC   P20164;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HEMORRHAGIC METALLOPROTEINASE HR1B (EC 3.4.24.52) (TRIMERELYSIN I).
OS   Trimeresurus flavoviridis (Habu).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Viperidae; Crotalinae; Trimeresurus.
OX   NCBI_TaxID=88087;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Venom;
RX   MEDLINE=90375459; PubMed=2398046;
RA   Takeya H., Oda K., Miyata T., Omori-Satoh T., Iwanaga S.;
RT   "The complete amino acid sequence of the high molecular mass
RT   hemorrhagic protein HR1B isolated from the venom of Trimeresurus
RT   flavoviridis.";
RL   J. Biol. Chem. 265:16068-16073(1990).
CC   -!- FUNCTION: THIS PROTEIN IS A ZINC PROTEASE FROM SNAKE VENOM THAT
CC       ACTS IN HEMORRHAGE.
CC   -!- FUNCTION: INHIBITS FIBRINOGEN INTERACTION WITH PLATELET RECEPTORS
CC       EXPRESSED ON GLYCOPROTEIN IIB-IIIA COMPLEX. ACT BY BINDING TO THE
CC       GLYCOPROTEIN IIB-IIIA RECEPTOR ON THE PLATELET SURFACE AND INHIBIT
CC       AGGREGATION INDUCED BY ADP, THROMBIN, PLATELET-ACTIVATING FACTOR
CC       AND COLLAGEN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: CLEAVAGE OF ONLY TWO BONDS: 10-HIS-|-LEU-11
CC       AND 14-ALA-|-LEU-15, IN THE INSULIN B CHAIN.
CC   -!- COFACTOR: BINDS ONE ZINC ION.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS A DISINTEGRIN DOMAIN.
DR   PIR; A37877; A37877.
DR   HSSP; P18619; 1FVL.
DR   MEROPS; M12.154; -.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF00200; disintegrin; 1.
DR   PRINTS; PR00289; DISINTEGRIN.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hydrolase; Metalloprotease; Zinc; Blood coagulation; Platelet;
KW   Cell adhesion; Venom; Zymogen; Glycoprotein.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   DOMAIN        1    203       METALLOPROTEASE.
FT   DOMAIN      204    300       DISINTEGRIN.
FT   DOMAIN      301    416       CYS-RICH.
FT   METAL       143    143       ZINC (CATALYTIC) (PROBABLE).
FT   ACT_SITE    144    144       BY SIMILARITY.
FT   METAL       147    147       ZINC (CATALYTIC) (PROBABLE).
FT   METAL       153    153       ZINC (CATALYTIC) (PROBABLE).
FT   DISULFID    118    198       BY SIMILARITY.
FT   DISULFID    158    165       BY SIMILARITY.
FT   DISULFID    225    243       BY SIMILARITY.
FT   DISULFID    227    238       BY SIMILARITY.
FT   DISULFID    237    260       BY SIMILARITY.
FT   DISULFID    251    257       BY SIMILARITY.
FT   DISULFID    256    282       BY SIMILARITY.
FT   DISULFID    269    289       BY SIMILARITY.
FT   CARBOHYD     73     73       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    181    181       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    327    327       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    380    380       N-LINKED (GLCNAC. . .).
SQ   SEQUENCE   416 AA;  46864 MW;  663FECFEB6186915 CRC64;

alignment_scores:
            Quality:  610.00              Length:    443
              Ratio:  2.302                Gaps:     16
 Percent Similarity:  59.819   Percent Identity: 31.828

alignment_block:
 US-09-634-252A-2 x HR1B_TRIFL   ..
```

```
Align seg 1/1  to: HR1B_TRIFL  from: 1  to: 416

 895 AAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCA 944
     :::|||:::::::|||   ||||||   ||||||              |||||
   7 ArgTyrIleLysLeuAlaIleValValAspHisGlyIleValThrLysHi 23

 945 TCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACC 994
     |:::::::::                :::::::::      ::::::||||||
  23 sHisGlyAsnLeuLysLysIleArgLysTrpIleTyrGlnLeuValAsnT 40

 995 TTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTG 1044
       :::::::::||||||:::   :::||||||      |||:::||||||
  40 hrIleAsnAsnIleTyrArg...SerLeuAsnIleLeuValAlaLeuVal 55

1045 GCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCC 1094
       :::|||   |||:::::::::::::::|||   :::   :::
  56 TyrLeuGluIleTrpSerLysGlnAsnLysIleThrValGlnSerAlaSe 72

1095 TGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATT......A 1138
               |||      |||:::   :::|||:::   :::      |
  72 rAsnValThrLeuAspLeuPheGlyAspTrpArgGluSerValLeuLeuL 89

 139 AGCAGCATGCT...GATGCTGTGCACCTCATCTCGCGGGTGACATTTCAC 1185
     |||||:::::::   |||:::::::::|||:::::::   :::   |||
  89 ysGlnArgSerHisAspCysAlaGlnLeuLeuThrThrIleAspPheAsp 105

1186 TATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAG 1235
                :::   :::|||   :::::::::|||:::      ||
 106 GlyProThrIleGlyLysAlaTyrThrAlaSerMetCysAspProLysAr 122

1236 AGGAGTTGGT...GTGAATGAGTATGGT...CTTCCAATGGCAGTGGCAC 1279
     |:::||||||   |||:::::|||:::   :::   ::::::||||||
 122 gSerValGlyIleValGlnAspTyrSerProIleAsnLeuValValAlaV 139

1280 AAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCT 1329
       :::::::::::::::::::::::::|||||||||||||      :::
 139 alIleMetThrHisGluMetGlyHisAsnLeuGlyIleProHisAspGly 155

1330 TCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGC..... 1374
     :::|||              |||   |||           ||||||
 156 AsnSer............CysThrCys............GlyGlyPhePr 164

1375 .TGCATCATGGAGGAAACAGGGGTGTCCCAT......TCTCGAAAATTTT 1417
      ||||||||||:::          :::|||         |||:::   ||||
 164 oCysIleMetSerProMet...IleSerAspProProSerGluLeuPheS 180

1418 CAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGA 1467
     ||:::||||||         |||:::   ||||||
 180 erAsnCysSerLysAlaTyrTyrGlnThrPheLeuThrAspHisLysPro 196

1468 GCCTGCCTTTTCAACAGGCCAACAAAG......CTATTTGAGCCCACGGA 1511
       |||:::::::|||   |||:::|||      :::   :::|||
 197 GlnCysIleLeuAsnAlaProSerLysThrAspIleValSerProProVa 213

1512 ATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTC 1561
      |||||||||      :::||||||||||||||||||||||||||||||
 213 lCysGlyAsnGluLeuLeuGluAlaGlyGluGluCysAspCysGlySerP 230

1562 ATGTGGAATGCTATGGATTATGCTGT......AAGAAATGTTCCCTCTCC 1605
          :::|||         ||||||         :::|||:::|||
 230 roGluAsnCysGlnTyrGlnCysCysAspAlaAlaSerCysLysLeuHis 246

1606 AACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCT 1655
     :::   :::   |||:::::::|||   ||||||:::      :::|||
 247 SerTrpValLysCysGluSerGlyGluCysCysAsp......GlnCysAr 261

1656 TTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATA 1705
      |||:::      |||   |||||||||   |||   :::||||||||||
 261 gPheArgThrAlaGlyThrGluCysArgAlaAlaGluSerGluCysAspI 278

1706 TTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCAT 1755
     ||   |||   |||||||||:::|||:::::::||||||   :::
 278 leProGluSerCysThrGlyGlnSerAlaAspCysProThrAspArgPhe 294

1756 AAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGG 1805
        :::::::|||      |||      |||:::|||   ||||||||||
 295 HisArgAsnGlyGlnProCysLeuTyrAsnHisGlyTyrCysTyrAsnGl 311

1806 CGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGG 1855
     |:::|||                |||||||   :::::::::|||:::::::|
 311 yLysCysProIleMetPheTyrGlnCysTyrPheLeuPheGlySerAsnA 328

1856 CTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACT 1905
     ||:::   ::::::      |||:::::::::   |||   :::|||
 328 laThrValAlaGluAspAspCysPheAsnAsn...AsnLysLysGlyAsp 343

1906 GAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAA 1955
     :::          |||   |||:::::::::::::::|||   |||:::::
 344 LysTyrPheTyrCysArgLysGluAsnGluLysTyrIleProCysAlaGl 360

1956 ACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTC 2005
     ::::||||||   ||||||   |||:::|||   |||
 360 nGluAspValLysCysGlyArgLeuPheCysAspAsn............. 372

2006 CACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCAT 2055

 372 ..................................................  372

2056 CAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGA 2105
            :::          |||          |||          :::::::||
 373 ......LysLysTyrProCys.........HisTyrAsnTyrSerGluAs 384

2106 TACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGA 2155
     |   |||   |||   |||:::   ||||||   |||:::   :::   :
 384 pLeuAspPheGlyMetValAspHisGlyThrLysCysAlaAspGlyLysV 401

2156 TGTGTTTAGATCGGAAGTGCCTACAAATT 2184
     ::|||   :::|||:::|||:::::::::
 401 alCysSerAsnArgGlnCysValAspVal 410

seq_name: SwissProt_39:HRTE_CROAT

seq_documentation_block:
ID   HRTE_CROAT     STANDARD;      PRT;   478 AA.
AC   P34182;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HEMORRHAGIC METALLOPROTEINASE HT-E PRECURSOR (EC 3.4.24.44)
DE   (ATROLYSIN E) (HEMORRHAGIC TOXIN E).
OS   Crotalus atrox (Western diamondback rattlesnake).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Viperidae; Crotalinae; Crotalus.
OX   NCBI_TaxID=8730;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Venom;
RX   MEDLINE=92329442; PubMed=1378300;
RA   Hite L.A., Shannon J.D., Bjarnason J.B., Fox J.W.;
RT   "Sequence of a cDNA clone encoding the zinc metalloproteinase
RT   hemorrhagic toxin e from Crotalus atrox: evidence for signal,
RT   zymogen, and disintegrin-like structures.";
RL   Biochemistry 31:6203-6211(1992).
CC   -!- FUNCTION: THIS PROTEIN IS A ZINC PROTEASE FROM SNAKE VENOM THAT
CC       ACTS IN HEMORRHAGE.
CC   -!- FUNCTION: INHIBITS FIBRINOGEN INTERACTION WITH PLATELET RECEPTORS
CC       EXPRESSED ON GLYCOPROTEIN IIB-IIIA COMPLEX. ACT BY BINDING TO THE
CC       GLYCOPROTEIN IIB-IIIA RECEPTOR ON THE PLATELET SURFACE AND INHIBIT
CC       AGGREGATION INDUCED BY ADP, THROMBIN, PLATELET-ACTIVATING FACTOR
CC       AND COLLAGEN (BY SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: CLEAVAGE OF 3-ASN-|-GLN-4, 9-SER-|-HIS-10 AND
CC       14-ALA-LEU-15 BONDS IN INSULIN B CHAIN AND 14-TYR-|-GLN1-5 AND
CC       8-ALA-|-SER-9 IN A CHAIN. CLEAVES TYPE IV COLLAGEN AT 258-ALA-|-
CC       GLN-259 IN ALPHA-1-(IV) AND AT 191-GLY-|-LEU-192 IN ALPHA-2-(IV).
CC   -!- COFACTOR: BINDS ONE ZINC ION.
```

```
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS A DISINTEGRIN DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M89784; AAB00731.1; -.
DR   PIR; A43296; A43296.
DR   HSSP; P18619; 1FVL.
DR   MEROPS; M12.145; -.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF00200; disintegrin; 1.
DR   PRINTS; PR00289; DISINTEGRIN.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hydrolase; Metalloprotease; Zinc; Blood coagulation; Platelet;
KW   Cell adhesion; Venom; Zymogen; Signal.
FT   SIGNAL       1    18       POTENTIAL.
FT   PROPEP      19   187       POTENTIAL.
FT   CHAIN      188   389       HEMORRHAGIC METALLOPROTEINASE HT-E.
FT   CHAIN      390   478       DISINTEGRIN (POTENTIAL).
FT   METAL      329   329       ZINC (CATALYTIC) (PROBABLE).
FT   ACT_SITE   330   330       BY SIMILARITY.
FT   METAL      333   333       ZINC (CATALYTIC) (PROBABLE).
FT   METAL      339   339       ZINC (CATALYTIC) (PROBABLE).
FT   DISULFID   304   384       BY SIMILARITY.
FT   DISULFID   344   351       BY SIMILARITY.
SQ   SEQUENCE   478 AA;  53637 MW;  02B2724F38D9D686 CRC64;


alignment_scores:
            Quality:  498.50              Length:    515
              Ratio:  1.656                 Gaps:     25
Percent Similarity:  58.447   Percent Identity: 28.738

Alignment_block:
US-09-634-252A-2 x HRTE_CROAT   ..

Align seg 1/1  to: HRTE_CROAT  from: 1  to: 478

    283 CAGAATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAAT 332
        |||:::||||||      ::::::   :::|||:::   :::   || ::
     17 GlnGlySerSerIleIleLeuGluSerGlyAsnValAsnAspTyrGluVa 33

    333 CACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGAAAGCC 382
        :      |||   ::::::
     33 lIleTyrProArgLysValThrAlaLeu..................... 42

    383 CTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAG 432
                           |||:::      |||   |||::::::
     43 .................ProLysGlyAlaValGlnProLysTyrGluAsp 53

    433 GCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAA 482
        ::::::                ::::::::::::::::::   |||:::
     54 ThrMet..............GlnTyrGluLeuLysValAsnGlyGluPr 65

    483 ATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATG 532
            :::|||   |||      |||:::||||||:::|||:::||||||
     65 oValValLeuHisLeuGluLysAsnLysGlyLeuPheSerLysAspTyrS 82

    533 TGGAGATTCACTACGAA...AATGGGAAACCACAGTACTCTAAGGGTGGA 579
          |||   ||||||:::   :::|||:::         ::::::   :::
     82 erGluThrHisTyrSerPheAspGlyArgLysIleThrThrAsnProSer 98

    580 ......GAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTC 623
              :::||||||||||||||||||   |||::::::      |||||
     99 ValGluAspHisCysTyrTyrHisGlyArgIleGluAsnAspAlaAspSe 115

    624 CAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATG 673
        |   :::::::::|||:::||||||||||||   |||   |||:::   :
    115 rThrAlaSerIleSerAlaCysAsnGlyLeuLysGlyHisPheLysLeuG 132

    674 ATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAA 723
        ::      :::|||:::||||||||||||:::|||
    132 lnGlyGluMetTyrLeuIleGluProLeuLysLeu.............. 143

    724 AGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCAGGACAGTATTC 773
        |||   ::::::   |||   :::   |||
    144 SerAspSerGluAlaHisAlaValPheLys..................... 153

    774 TAAGCAAATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTC 823
              :::||||||      :::|||:::   |||:::   |||   :
    154 .......LeuLysAsn......ValGluLysGluAspGluAlaProLysM 166

    824 TCTCTGAATTA......CAGTGG.................TTGAAAAGA 849
        ::      :::      :::|||                    |||
    166 etCysGlyValThrGlnAsnTrpGluSerTyrGluProIleLysLysAla 182

    850 AGGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATA 899
                     :::||||||               |||:::::::::||
    183 SerAspLeuAsnLeuAsnPro..............GluHisGlnArgTy 194

    900 TTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCT 949
        |:::||||||:::||||||   ||||||      |||   |||::::::::
    194 rValGluLeuPheIleValValAspHisGlyMetTyrThrLysTyrAsnG 211

    950 CTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTG 999
        :::::              ::::::   :::   ::::::|||||:::::
    211 lyAspSerAspLysIleArgGlnArgValHisGlnMetValAsnIleMet 227

   1000 GATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGT 1049
          :::   |||          :::         ::::::|||:::::::::
    228 LysGluSerTyr...ThrTyrMetTyrIleAspIleLeuLeuAlaGlyIl 243

   1050 AGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGC 1099
        :|||   |||::::::   |||   |||::::::      .   :
    243 eGluIleTrpSerAsnGlyAspLeuIleAsnValGlnProAlaSerProA 260

   1100 AGATGCTCCATGAGTTCTCAAAATACCGGCAG............CGCATT 1137
        ::   |||:::::::|||:::::::::|||:::             :::
    260 snThrLeuAsnSerPheGlyGluTrpArgGluThrAspLeuLeuLysArg 276

   1138 AAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTA 1187
        |||:::|||   |||   ::::::|||::::::   ::::::|||
    277 LysSerHis...AspAsnAlaGlnLeuLeuThrSerIleAlaPheAspGl 292

   1188 TAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAG 1237
          :::      :::   :::|||:::||||||:::|||:::      |||:
    292 uGlnIleIleGlyArgAlaTyrIleGlyGlyIleCysAspProLysArgS 309

   1238 GAGTTGGTGTG......AATGAGTATGGTCTTCCAATGGCAGTGGCACAA 1281
        :::::||||||      ::::::      :::   :::   ||||||
    309 erThrGlyValValGlnAspHisSerGluIleAsnLeuArgValAlaVal 325

   1282 GTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTC 1331
        ::::::::::::::::|||:::::::||||||||||||:::   :::   ::
    326 ThrMetThrHisGluLeuGlyHisAsnLeuGlyIleHisHisAspThrAs 342

   1332 TAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGC......T 1375
        :|||            |||:::|||            ||||||      |
    342 pSer............CysSerCys............GlyGlyTyrSerC 351

   1376 GCATCATGGAGGAAACA...GGGGTGTCCCATTCTCGAAAATTTTCAAAG 1422
```

```
     |||||||||:::   :::   :::   :::   |||:::   ||||||
 351 ysIleMetSerProValIleSerAspGluProSerLysTyrPheSerAsp 367

1423 TGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTG 1472
     ||||||   ::::::      :::|||:::::::::            ||
 368 CysSerTyrIleGlnCysTrpGluPheIleMetAsnGlnLysProGlnCy 384

1473 CCTTTTCAACAGGCCA......ACAAAGCTATTTGAGCCCACGGAATGTG 1516
     |:::::::::::::|||      |||        :::           |
 384 sIleLeuLysLysProLeuArgThrAspThrValSerThrProValSerG 401

1517 GAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTG 1566
     |||||      :::||||||||||   ||||||||||||||||      :::
 401 lyAsnGluLeuLeuGluAlaGlyIleGluCysAspCysGly...SerLeu 416

1567 GAA.........TGCTATGGATTATGCTGTAAGAAATGTTCCCTCTCCAA 1607
     |||         ||||||:::      ||||||         :::
 417 GluAsnProCysCysTyrAlaThrThrCysLys.........MetArgPr 430

1608 CGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTT 1657
      |||::::::|||::::::|||   ||||||:::      :::|||   |
 430 oGlySerGlnCysAlaGluGlyLeuCysCysAsp......GlnCysArgP 445

1658 TTCAGCCACGAGGGTATGAATGCCGG...GATGCTGTGAACGAGTGTGAT 1704
     ||:::   :::|||      ||||||   :::   |||::::::   |||
 445 heMetLysLysGlyThrValCysArgValSerMetValAspArgAsnAsp 461

1705 ATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAAT 1749
        |||      |||||||||:::|||::::::||||||   |||
 462 AspThr......CysThrGlyGlnSerAlaAspCysProArgAsn 474
```

seq_name: SwissProt_39:DISA_TRIGA

seq_documentation_block:

```
ID   DISA_TRIGA     STANDARD;      PRT;   480 AA.
AC   P15503;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   PUTATIVE VENOM METALLOPROTEINASE PRECURSOR (EC 3.4.24.-) [CONTAINS:
DE   DISINTEGRIN TRIGRAMIN ALPHA (PLATELET AGGREGATION ACTIVATION
DE   INHIBITOR)].
OS   Trimeresurus gramineus (Indian green tree viper) (Green habu snake).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Viperidae; Crotalinae; Trimeresurus.
OX   NCBI_TaxID=8767;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Venom gland;
RX   MEDLINE=90332429; PubMed=2377470;
RA   Neeper M.P., Jacobson M.A.;
RT   "Sequence of a cDNA encoding the platelet aggregation inhibitor
RT   trigramin.";
RL   Nucleic Acids Res. 18:4255-4255(1990).
RN   [2]
RP   SEQUENCE OF 408-479.
RX   MEDLINE=89229063; PubMed=2653425;
RA   Huang T.-F., Holt J.C., Kirby E.P., Niewiarowski S.;
RT   "Trigramin: primary structure and its inhibition of von Willebrand
RT   factor binding to glycoprotein IIb/IIIa complex on human platelets.";
RL   Biochemistry 28:661-666(1989).
RN   [3]
RP   SEQUENCE OF 408-479.
RC   TISSUE=Venom;
RX   MEDLINE=90207217; PubMed=2320569;
RA   Dennis M.S., Henzel W.J., Pitti R.M., Lipari M.T., Napier M.A.,
RA   Deisher T.A., Bunting S., Lazarus R.A.;
RT   "Platelet glycoprotein IIb-IIIa protein antagonists from snake
RT   venoms: evidence for a family of platelet-aggregation inhibitors.";
RL   Proc. Natl. Acad. Sci. U.S.A. 87:2471-2475(1990).
RN   [4]
RP   SEQUENCE OF 408-419.
RX   MEDLINE=88058981; PubMed=3680247;
RA   Huang T.-F., Holt J.C., Lukasiewicz H., Niewiarowski S.;
RT   "Trigramin. A low molecular weight peptide inhibiting fibrinogen
RT   interaction with platelet receptors expressed on glycoprotein
RT   IIb-IIIa complex.";
RL   J. Biol. Chem. 262:16157-16163(1987).
CC   -!- FUNCTION: THIS PROTEIN IS A ZINC PROTEASE FROM SNAKE VENOM THAT
CC       ACTS IN HEMORRHAGE (BY SIMILARITY).
CC   -!- FUNCTION: TRIGRAMIN INHIBITS FIBRINOGEN INTERACTION WITH PLATELET
CC       RECEPTORS EXPRESSED ON GLYCOPROTEIN IIB-IIIA COMPLEX. ACT BY
CC       BINDING TO THE GLYCOPROTEIN IIB-IIIA RECEPTOR ON THE PLATELET
CC       SURFACE AND INHIBITS AGGREGATION INDUCED BY ADP, THROMBIN,
CC       PLATELET-ACTIVATING FACTOR AND COLLAGEN.
CC   -!- COFACTOR: BINDS ONE ZINC ION.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: IN THE C-TERMINAL REGION; BELONGS TO THE DISINTEGRIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X51530; CAA35910.1; -.
DR   PIR; A29784; A29784.
DR   PIR; A30065; A30065.
DR   HSSP; P18619; 1FVL.
DR   MEROPS; M12.163; -.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF00200; disintegrin; 1.
DR   PRINTS; PR00289; DISINTEGRIN.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hydrolase; Metalloprotease; Zinc; Blood coagulation; Platelet;
KW   Cell adhesion; Venom; Zymogen; Signal.
FT   SIGNAL        1      ?       POTENTIAL.
FT   PROPEP        ?    190       POTENTIAL.
FT   CHAIN       191    407       PUTATIVE ZINC METALLOPROTEINASE.
FT   CHAIN       408    480       TRIGRAMIN.
FT   METAL       333    333       ZINC (CATALYTIC) (PROBABLE).
FT   ACT_SITE    334    334       BY SIMILARITY.
FT   METAL       337    337       ZINC (CATALYTIC) (PROBABLE).
FT   METAL       343    343       ZINC (CATALYTIC) (PROBABLE).
FT   DISULFID    308    386       BY SIMILARITY.
FT   DISULFID    348    353       BY SIMILARITY.
FT   DISULFID    413    428       BY SIMILARITY.
FT   DISULFID    415    423       BY SIMILARITY.
FT   DISULFID    422    445       BY SIMILARITY.
FT   DISULFID    436    442       BY SIMILARITY.
FT   DISULFID    441    466       BY SIMILARITY.
FT   DISULFID    454    473       BY SIMILARITY.
FT   SITE        458    460       CELL ATTACHMENT SITE.
FT   CARBOHYD    279    279       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   480 AA;  53494 MW;  601415FF3A4E0D50 CRC64;
```

alignment_scores:

```
           Quality:  494.50              Length:    510
             Ratio:  1.665                 Gaps:     21
 Percent Similarity:  58.235   Percent Identity: 28.431
```

alignment_block:

```
US-09-634-252A-2 x DISA_TRIGA  ..

Align seg 1/1  to: DISA_TRIGA  from: 1  to: 480

    283 CAGAATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAAT 332
        |||:::|||||      ::::::   :::|||      :::   |||::
     17 GlnGlySerSerIleIleLeuGluSerGlyAsnLeuAsnAspTyrGluVa 33

    333 CACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGAAAGCC 382
        ::::   |||:::::::::
     33 lValTyrProGluLysValThrAlaLeu..................... 42

    383 CTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAG 432
                        |||:::      |||||||||:::::
     43 .................ProLysGlyAlaValGlnGlnLysTyrGluAsp 53

    433 GCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAA 482
        |||:::                 ::::::::::::::::::   |||:::
     54 AlaMet..............GlnTyrGluPheLysValAsnGlyGluPr 65

    483 ATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATG 532
          :::|||   |||      |||:::||||||:::|||:::||||||
     65 oValValLeuHisLeuGluLysAsnLysGlyLeuPheSerGluAspTyrS 82

    533 TGGAGATTCACTACGAA...AATGGGAAA......CCACAGTACTCTAAG 573
        ||||||||||||:::   :::|||:::             |||   :::
     82 erGluIleHisTyrSerProAspGlyArgGluIleThrAlaTyrProSer 98

    574 GGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTC 623
          :::||||||||||||||||||   |||:::::      |||||
     99 ValGluAspHisCysTyrTyrHisGlyArgIleGluAsnAspAlaAspSe 115

    624 CAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATG 673
        |   :::::::::|||:::|||:::||||||   |||   |||:::   :
    115 rThrAlaSerIleSerAlaCysAspGlyLeuLysGlyHisPheLysLeuG 132

    674 ATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAA 723
        ::      :::|||:::||||||||||||||||||
    132 lnGlyGluMetTyrLeuIleGluProLeuGluLeu.............. 143

    724 AGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCAGGACAGTATTC 773
        |||   ::::::   |||   :::
    144 SerAspSerGluAlaHisAlaVal........................ 151

    774 TAAGCAAATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTC 823
              :::|||      ::::::|||:::   |||:::  .|||   :
    152 ..........PheLysTyrGluAsnValGluLysGluAspGluProProLysM 166

    824 TCTCTGAATTA......CAGTGG......TTGAAAAGAAGGAAGAGAGCA 861
        ::      :::      :::|||          :::      |||:::|||
    166 etCysGlyValThrGlnAsnTrpGluSerTyrGluSerThrLysLysAla 182

    862 ............GTGAATCCATCACGTGGTATATTTGAAGAAATGAAATA 899
                  |||:::|||::::::      |||   :::   :::||
    183 SerGlnLeuAsnValThrProGluGlnGlnArgPheProGln...ArgTy 198

    900 TTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCT 949
        |::::::|||   |||      ||||||      |||   |||:::   :
    198 rIleLysLeuGlyIlePheValAspHisGlyMetTyrThrLysTyrSerG 215

    950 CTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTG 999
        :::::      :::   ::::::   :::   :::::::::|||   :::
    215 lyAsnSerGluArgIleThrLysArgValHisGlnMetIleAsnAsnIle 231

   1000 GATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGT 1049
        :::   :::   :::      ||||||      :::::|||     :::::
    232 AsnMetMetCysArg...AlaLeuAsnIleValThrThrLeuSerValLe 247

   1050 AGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGC 1099
        :|||   |||:::|||||||||   |||   :::   ::::::
    247 uGluIleTrpSerGluLysAspLeuIleThrValGlnAlaSerAlaProT 264

   1100 AGATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATT.....AAG 1140

            |||      |||:::   :::|||:::   :::         :::
    264 hrThrLeuThrLeuPheGlyAlaTrpArgGluThrValLeuLeuAsnArg 280

   1141 CAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAA 1190
                |||   ::::::|||::::::   :::   |||:::   ::
    281 ThrSerHisAspHisAlaGlnLeuLeuThrAlaThrIlePheAsnGlyAs 297

   1191 GAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAG 1240
        :      :::   :::      ||||||:::|||:::      |||:::|
    297 nValIleGlyArgAlaProValGlyGlyMetCysAspProLysArgSerV 314

   1241 TTGGTGTG......AATGAGTATGGTCTTCCAATGGCAGTGGCACAAGTA 1284
        ||::::::      ::::::   ::::::   ::::::||||||   :::
    314 alAlaIleValArgAspHisAsnAlaIleValPheValValAlaValThr 330

   1285 TTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAG 1334
        :::::::::::::::::::::||||||||:::::   :::
    331 MetThrHisGluMetGlyHisAsnLeuGlyMetHisHisAsp........ 344

   1335 CAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCATGG 1384
            :::   ||||||:::|||:::                ||||||||:
    345 ....GluAspLysCysAsnCysAsnThr............CysIleMetS 356

   1385 AGGAAACA...GGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATT 1431
        ::::::::   :::   :::   |||:::   ||||||:::||||||
    356 erLysValLeuSerArgGlnProSerLysTyrPheSerGluCysSerLys 372

   1432 TTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAA 1481
              |||:::   ||||||   :::   :::      |||:::::||
    373 AspTyrTyrGlnThrPheLeuThrAsnHisAsnProGlnCysIleLeuAs 389

   1482 CAGGCCA......ACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGAT 1525
        |   |||      |||         :::             ||||||
    389 nAlaProLeuArgThrAspThrValSerThrProValSerGlyAsnGluL 406

   1526 ACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTAT 1575
          :::||||||||||||:::||||||||||||      :::
    406 euLeuGluAlaGlyGluAspCysAspCysGlySerProAla......... 419

   1576 GGATTATGCTGT......AAGAAATGTTCCCTCTCCAACGGGGCTCACTG 1619
        :::   ||||||            |||:::|||      ||||||:::||
    420 AsnProCysCysAspAlaAlaThrCysLysLeuIleProGlyAlaGlnCy 436

   1620 CAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAG 1669
        |:::::|||   ||||||:::      :::|||   |||      :::|
    436 sGlyGluGlyLeuCysCysAsp......GlnCysSerPheIleGluGluG 451

   1670 GGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGT 1719
        ||      ||||||   |||   ::::::   |||:::   :::||||||
    451 lyThrValCysArgIleAlaArgGlyAsp...AspLeuAspAspTyrCys 466

   1720 ACTGGAGACTCTGGTCAGTGCCCACCAAAT 1749
        :::|||   |||:::   ||||||   |||
    467 AsnGlyArgSerAlaGlyCysProArgAsn 476

seq_name: SwissProt_39:DISR_AGKRH

seq_documentation_block:
ID   DISR_AGKRH     STANDARD;      PRT;   478 AA.
AC   P30403;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   HEMORRHAGIC PROTEIN-RHODOSTOMIN PRECURSOR (EC 3.4.24.-) (RHO)
DE   [CONTAINS: DISINTEGRIN RHODOSTOMIN].
GN   RHOD.
OS   Agkistrodon rhodostoma (Malayan pit viper) (Calloselasma rhodostoma).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Viperidae; Crotalinae; Calloselasma.
OX   NCBI_TaxID=8717;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   TISSUE=Venom gland;
RX   MEDLINE=93277962; PubMed=7916635;
RA   Au L.-C.;
RT   "Nucleotide sequence of a full-length cDNA encoding a common
RT   precursor of platelet aggregation inhibitor and hemorrhagic protein
RT   from Calloselasma rhodostoma venom.";
RL   Biochim. Biophys. Acta 1173:243-245(1993).
RN   [2]
RP   SEQUENCE OF 77-478 FROM N.A.
RC   TISSUE=Venom gland;
RX   MEDLINE=92095959; PubMed=1755841;
RA   Au L.-C., Huang Y.-B., Huang T.-F., Teh G.-W., Lin H.-H., Choo K.-B.;
RT   "A common precursor for a putative hemorrhagic protein and
RT   rhodostomin, a platelet aggregation inhibitor of the venom of
RT   Calloselasma rhodostoma: molecular cloning and sequence analysis.";
RL   Biochem. Biophys. Res. Commun. 181:585-593(1991).
RN   [3]
RP   SEQUENCE OF 408-475 FROM N.A.
RX   MEDLINE=93135775; PubMed=7916592;
RA   Chang H.H., Hu S.T., Huang T.-F., Chen S.H., Lee Y.H., Lo S.J.;
     "Rhodostomin, an RGD-containing peptide expressed from a synthetic
     gene in Escherichia coli, facilitates the attachment of human
     hepatoma cells.";
     Biochem. Biophys. Res. Commun. 190:242-249(1993).
     [4]
RP   SEQUENCE OF 408-475.
RX   MEDLINE=91046142; PubMed=2236100;
RA   Gould R.J., Polokoff M.A., Friedman P.A., Huang T.-F., Holt J.C.,
RA   Cook J.J., Niecviarowski S.;
RT   "Disintegrins: a family of integrin inhibitory proteins from viper
RT   venoms.";
RL   Proc. Soc. Exp. Biol. Med. 195:168-171(1990).
CC   -!- FUNCTION: THIS PROTEIN IS A ZINC PROTEASE FROM SNAKE VENOM THAT
CC       ACTS IN HEMORRHAGE.
CC   -!- FUNCTION: INHIBITS FIBRINOGEN INTERACTION WITH PLATELET RECEPTORS
CC       EXPRESSED ON GLYCOPROTEIN IIB-IIIA COMPLEX. ACT BY BINDING TO THE
CC       GLYCOPROTEIN IIB-IIIA RECEPTOR ON THE PLATELET SURFACE AND INHIBIT
CC       AGGREGATION INDUCED BY ADP, THROMBIN, PLATELET-ACTIVATING FACTOR
CC       AND COLLAGEN.
CC   -!- COFACTOR: BINDS ONE ZINC ION.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: IN THE C-TERMINAL REGION; BELONGS TO THE DISINTEGRIN
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L08780; AAA49196.1; -.
DR   PIR; JQ1301; JQ1301.
DR   PIR; S33792; S33792.
DR   HSSP; P17494; 1KST.
DR   MEROPS; M12.161; -.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF00200; disintegrin; 1.
DR   PRINTS; PR00289; DISINTEGRIN.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hydrolase; Metalloprotease; Zinc; Blood coagulation; Platelet;
KW   Cell adhesion; Venom; Zymogen; Glycoprotein; Signal.
FT   SIGNAL        1     18       POTENTIAL.
FT   PROPEP       19    188       POTENTIAL.
FT   CHAIN       189    391       HEMORRHAGIC PROTEIN.
FT   PROPEP      392    407
FT   CHAIN       408    475       RHODOSTOMIN.
FT   PROPEP      476    478
FT   METAL       330    330       ZINC (CATALYTIC) (PROBABLE).
FT   ACT_SITE    331    331       BY SIMILARITY.
FT   METAL       334    334       ZINC (CATALYTIC) (PROBABLE).
FT   METAL       340    340       ZINC (CATALYTIC) (PROBABLE).
FT   DISULFID    305    386       BY SIMILARITY.
FT   DISULFID    345    353       BY SIMILARITY.
FT   DISULFID    411    426       BY SIMILARITY.
FT   DISULFID    413    421       BY SIMILARITY.
FT   DISULFID    420    443       BY SIMILARITY.
FT   DISULFID    434    440       BY SIMILARITY.
FT   DISULFID    439    464       BY SIMILARITY.
FT   DISULFID    452    471       BY SIMILARITY.
FT   SITE        456    458       CELL ATTACHMENT SITE.
FT   CARBOHYD    279    279       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    369    369       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   478 AA;  54006 MW;  6490A2B171D3A830 CRC64;


alignment_scores:
            Quality:  475.50              Length:    513
              Ratio:  1.623                Gaps:     22
 Percent Similarity:  57.115   Percent Identity: 26.901

alignment_block:
 US-09-634-252A-2 x DISR_AGKRH   ..

 Align seg 1/1  to: DISR_AGKRH  from: 1  to: 478

    283 CAGAATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAAT 332
        |||:::||||||      ::::::   :::|||:::   :::   ||||:
     17 GlnGlySerSerIleIleLeuGluSerGlyAsnValAsnAspTyrGluVa 33

    333 CACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGAAAGCC 382
        ::::   |||   :::::::|||
     33 lValTyrProArgLysValIleAlaLeu..................... 42

    383 CTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAG 432
                       :::::::::      |||||||||::::::
     43 ................SerGluGlyAlaAlaGlnGlnLysTyrGluAsp 53

    433 GCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAA 482
        ::::::                ::::::::::::::::::   ||||:::
     54 ThrMet..............GlnTyrGluPheLysValAsnGlyGluPr 65

    483 ATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATG 532
           :::|||   |||      |||:::||||||:::::::::||||||
     65 oValValLeuHisLeuGluLysAsnLysGlyLeuPheAlaLysAspTyrS 82

    533 TGGAGATTCACTACGAAAATGGGAAACCACAG.........TACTCTAAG 573
          |||   ||||||:::           :::         |||   :::
     82 erGluThrHisTyrSerProAspGlyThrArgIleThrThrTyrProSer 98

    574 GGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTC 623
           :::||||||||||||:::|||   |||:::::      |||||
     99 ValGluAspHisCysTyrTyrGlnGlyArgIleHisAsnAspAlaAspSe 115

    624 CAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATG 673
        |   :::::::::|||:::||||||||||||   |||  |||:::   :
    115 rThrAlaSerIleSerAlaCysAsnGlyLeuLysGlyHisPheLysLeuG 132

    674 ATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAA 723
        ::      :::|||:::||||||||:::::::|||      :::|||
    132 lnGlyGluThrTyrPheIleGluProMetLysLeuProAspSerGlu... 147

    724 AGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCAGGACAGTATTC 773
                     |||   :::
    148 ............AlaHisAlaVal........................ 151
```

```
 774 TAAGCAAATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTC 823.
          :::|||      ::::::|||:::   |||:::   |||   :
 152 .......PheLysTyrGluAsnIleGluLysGluAspGluSerProLysM 166

 824 TC.........TCTGAATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTG 864
     ::         :::|||   :::|||                 ::::::
 166 etCysGlyValThrGluThrAsnTrp.............GluSerAsp 177

 865 AATCCATCACGTGGTATA.................TTTGAAGAAATGAA 896
     :::|||   :::   :::                  |||      ::
 178 GluProIleLysLysValSerGlnLeuAsnLeuAsnHisGluIleLysAr 194

 897 ATATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATC 946
     ::::::::::::::::::|||   |||   :::          ||||||
 194 gHisValAspIleValValValValAspSerArgPheCysThrLysHisS 211

 947 GCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTT 996
       ::::::            ::::::|||:::   :::|||||||||
 211 erAsnAspLeuGluValIleArgLysPheValHisGluValValAsnAla 227

 997 GTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGC 1046
     :::   :::   ||||||      ::::::      :::   |||||
 228 IleIleGluSerTyrLys...TyrMetHisPheGlyIleSerLeuVal As 243

1047 TGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTG 1096
      :::||||||||||   :::   |||   |||:::::::      :::
 243 nLeuGluThrTrpCysAsnGlyAspLeuIleAsnValGlnGluAspSerT 260

1097 TGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGC......ATTAAG 1140
       :::   |||      |||:::|||:::|||:::          ||||||
 260 yrGluThrLeuLysAlaPheGlyLysTrpArgGluSerAspLeuIleLys 276

1141 ......CAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCA 1184
           :::|||:::::::|||      |||::::::   :::   |||
 277 HisValAsnHisSerAsnAlaGlnPheLeuMetAsp...MetLysPheIl 292

1185 CTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAA 1234
        :::      :::   :::|||:::   ::::::|||:::      |
 292 eLysAsnIleIleGlyLysAlaTyrLeuAspSerIleCysAspProGluA 309

1235 GAGGAGTTGGTGTG......AATGAGTATGGTCTTCCAATGGCAGTGGCA 1278
     ||:::||||||:::      |||   :::|||:::   :::   ||||||
 309 rgSerValGlyIleValGlnAsnTyrHisGlyIleThrLeuAsnValAla 325

1279 CAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACC 1328
       :::::::::::::::::::::::::||||||||:::::   :::
 326 AlaIleMetAlaHisGluMetGlyHisAsnLeuGlyValArgHisAspGl 342

1329 TTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCA 1378
      :::              |||   |||      |||   :::   ||||
 342 yGlu...........TyrCysThrCysTyrGlySer...SerGluCysI 354

1379 TCATGGAGGAAACAGGGGTGTCCCAT.............TCTCGAAAA 1413
     |||||:::           ||||||              |||:::
 354 leMetSer...........SerHisIleSerAspProProSerLysTyr 366

1414 TTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGG 1463
     ||||||:::||||||      ::::::      ::::::::::::   ::
 367 PheSerAsnCysSerTyrTyrGlnPheTrpLysTyrIleGluAsnGlnAs 383

1464 TGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAAT 1513
     :      |||:::::::|||:::|||   :::      :::
 383 nProGlnCysIleLeuAsnLysProLeuArgThrValSerIleProValS 400

1514 GTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCAT 1563
       ||||||   ::::::||||||||:::||||||||||||:::
 400 erGlyAsnGluHisLeuGluAlaGlyLysGluCysAspCysSer..... 414

1564 GTGGAATGCTATGGATTATGCTGT......AAGAAATGTTCCCTCTCCAA 1607
       :::      :::   ||||||           |||:::|||
 415 ...SerProGluAsnProCysCysAspAlaAlaThrCysLysLeuArgPr 430

1608 CGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTT 1657
```

```
          ||||||:::|||::::::|||   ||||||:::      :::|||   |
      430 oGlyAlaGlnCysGlyGluGlyLeuCysCysGlu......GlnCysLysP 445

     1658 TTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATT 1707
          ||:::      |||      ||||||      :::   :::   |||:::
      445 heSerArgAlaGlyLysIleCysArg......IleProArgGlyAspMet 459

     1708 ...ACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCA 1743
               :::   ||||||||||:::|||::::::||||||
      460 ProAspAspArgCysThrGlyGlnSerAlaAspCysPro 472
```

```
seq_name: SwissProt_39:HRTD_CROAT

seq_documentation_block:
ID   HRTD_CROAT     STANDARD;      PRT;   414 AA.
AC   P15167;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   HEMORRHAGIC METALLOPROTEINASE HT-D AND HT-C PRECURSOR (EC 3.4.24.42)
DE   (ATROLYSINS D AND C) (HEMORRHAGIC TOXINS C AND D).
OS   Crotalus atrox (Western diamondback rattlesnake).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Viperidae; Crotalinae; Crotalus.
OX   NCBI_TaxID=8730;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Venom;
RA   Hite L.A., Jia L.G., Bjarnason J.B., Fox J.W.;
RL   Submitted (SEP-1993) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 191-393.
RC   TISSUE=Venom;
RX   MEDLINE=89308543; PubMed=2745407;
RA   Shannon J.D., Baramova E.N., Bjarnason J.B., Fox J.W.;
RT   "Amino acid sequence of a Crotalus atrox venom metalloproteinase
RT   which cleaves type IV collagen and gelatin.";
RL   J. Biol. Chem. 264:11575-11583(1989).
RN   [3]
RP   X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX   MEDLINE=94359948; PubMed=8078901;
RA   Zhang D., Botos I., Gomis-Ruth F.X., Doll R., Blood C., Njoroge F.G.,
RA   Fox J.W., Bode W., Meyer E.F.;
RT   "Structural interaction of natural and synthetic inhibitors with the
RT   venom metalloproteinase, atrolysin C (form d).";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:8447-8451(1994).
CC   -!- FUNCTION: THIS PROTEIN IS A ZINC PROTEASE FROM SNAKE VENOM THAT
CC       ACTS IN HEMORRHAGE. IT CLEAVES TYPE IV COLLAGEN AND GELATIN.
CC   -!- CATALYTIC ACTIVITY: CLEAVAGE OF 5-HIS-|-LEU-6, 10-HIS-|-LEU-11,
CC       14-ALA-|-LEU-15, 16-TYR-|-LEU-17 AND 23-GLY-|-PHE-24 OF INSULIN B
CC       CHAIN. WITH SMALL MOLECULE SUBSTRATES PREFERS HYDROPHOBIC RESIDUE
CC       AT P2' AND SMALL RESIDUE SUCH AS ALA, GLY AT P1.
CC   -!- COFACTOR: BINDS ONE ZINC ION.
CC   -!- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF HT-D.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS A DISINTEGRIN DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U01237; AAA03352.1; -.
DR   PIR; A34166; HYRSAC.
DR   PDB; 1ATL; 15-OCT-95.
DR   PDB; 1DTH; 12-FEB-97.
DR   PDB; 1HTD; 15-SEP-95.
DR   MEROPS; M12.144; -.
```

```
DR   InterPro; IPR000130; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hydrolase; Metalloprotease; Zinc; Venom; Signal; 3D-structure.
FT   SIGNAL        1      ?       POTENTIAL.
FT   PROPEP        ?    190
FT   CHAIN       191    393       HEMORRHAGIC METALLOPROTEINASE HT-D.
FT   CHAIN       394    414       DISINTEGRIN (POTENTIAL).
FT   MOD_RES     191    191       PYRROLIDONE CARBOXYLIC ACID.
FT   METAL       333    333       ZINC (CATALYTIC).
FT   ACT_SITE    334    334
FT   METAL       337    337       ZINC (CATALYTIC).
FT   METAL       343    343       ZINC (CATALYTIC).
FT   DISULFID    308    388
FT   DISULFID    348    355
FT   VARIANT     191    191       MISSING (IN SOME CHAINS).
FT   VARIANT     371    371       D -> A (IN HT-C).
SQ   SEQUENCE   414 AA;  46845 MW;  0B41F26C9E450031 CRC64;

alignment_scores:
            Quality:  368.50              Length:    443
              Ratio:  1.396                 Gaps:     19
 Percent Similarity:  59.594   Percent Identity: 26.185

alignment_block:
 US-09-634-252A-2 x HRTD_CROAT   ..

 Align seg 1/1  to: HRTD_CROAT  from: 1  to: 414

     283 CAGAATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAAT 332
          |||:::||||||      ::::::   :::|||:::   :::   ||||:
      17 GlnGlySerSerIleIleLeuGluSerGlyAsnValAsnAspTyrGluVa 33

     333 CACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGAAAGCC 382
          ::::   |||   ::::::
      33 lValTyrProArgLysValThrAlaLeu..................... 42

     383 CTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAG 432
                              |||:::      |||   |||::::::
      43 .................ProLysGlyAlaValGlnProLysTyrGluAsp 53

     433 GCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAA 482
          |||:::                 ::::::::::::::::::   |||:::
      54 AlaMet..............GlnTyrGluLeuLysValAsnGlyGluPr 65

     483 ATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATG 532
              :::|||   |||      |||:::   |||:::|||:::||||||
      65 oValValLeuHisLeuGluLysAsnLysGluLeuPheSerLysAspTyrS 82

      33 TGGAGATTCACTACGAA...AATGGGAAACCACAGTACTCTAAGGGTGGA 579
           |||   ||||||:::   :::|||:::          ::::::   :::
      82 erGluThrHisTyrSerProAspGlyArgLysIleThrThrAsnProSer 98

     580 ......GAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTC 623
               :::||||||||||||:::|||   |||::::::      |||||
      99 ValGluAspHisCysTyrTyrArgGlyArgIleGluAsnAspAlaAspSe 115

     624 CAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATG 673
          |   ::::::::::|||:::||||||||||||   |||   |||:::   :
     115 rThrAlaSerIleSerAlaCysAsnGlyLeuLysGlyHisPheLysLeuG 132

     674 ATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAA 723
          ::      :::|||:::||||||||||||||||||
     132 lnGlyGluMetTyrLeuIleGluProLeuGluLeu.............. 143

     724 AGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCAGGACAGTATTC 773
          |||   ::::::   |||  :::   |||
     144 SerAspSerGluAlaHisAlaValPheLys..................... 153

     774 TAAGCAAATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTC 823
                 :::::::|||      :::|||:::   |||:::   |||   :
     154 .......LeuGluAsn......ValGluLysGluAspGluAlaProLysM 166

     824 TCTCTGAATTA......CAGTGG.................TTGAAAAGA 849
          ::      :::      :::|||                    |||
     166 etCysGlyValThrGlnAsnTrpGluSerTyrGluProIleLysLysAla 182

     850 AGGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATA 899
                      :::||||||::::::      :::   :::   :::||
     183 SerAspLeuAsnLeuAsnProAspGlnGlnAsnLeuProGln...ArgTy 198

     900 TTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCT 949
          |:::||||||::::::|||   ||||||:::::::::   |||::::::|
     198 rIleGluLeuValValValAlaAspHisArgValPheMetLysTyrAsnS 215

     950 CTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTG 999
          ||:::               ::::::   :::   ::::::||||||::::::
     215 erAspLeuAsnThrIleArgThrArgValHisGluIleValAsnPheIle 231

    1000 GATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGT 1049
          ::::::::::|||:::   :::||||||   :::|||   |||:::   ::
     232 AsnGlyPheTyrArg...SerLeuAsnIleHisValSerLeuThrAspLe 247

    1050 AGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGC 1099
          :|||   |||::::::::::|||||||||:::|||   :::          :
     247 uGluIleTrpSerAsnGluAspGlnIleAsnIleGlnSerAlaSerSerA 264

    1100 AGATGCTCCATGAGTTCTCAAAATACCGGCAG............CGCATT 1137
          ::   |||:::   |||::::::::::|||:::          :::
     264 spThrLeuAsnAlaPheAlaGluTrpArgGluThrAspLeuLeuAsnArg 280

    1138 AAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTA 1187
          |||:::|||   |||   ::::::|||::::::   :::   :::
     281 LysSerHis...AspAsnAlaGlnLeuLeuThrAlaIleGluLeuAspGl 296

    1188 TAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAG 1237
          :::      :::|||:::   :::|||   :::|||:::          :
     296 uGluThrLeuGlyLeuAlaProLeuGlyThrMetCysAspProLysLeuS 313

    1238 GAGTTGGTGTGAATGAGTATGGTCTTCCA......ATGGCAGTGGCACAA 1281
          :::::|||:::   :::         |||      :::   ::::::
     313 erIleGlyIleValGlnAspHisSerProIleAsnLeuLeuMetGlyVal 329

    1282 GTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTC 1331
          ::::::::::::::|||::::::||||||||||::::::   :::   ::
     330 ThrMetAlaHisGluLeuGlyHisAsnLeuGlyMetGluHisAspGlyLy 346

    1332 TAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCA 1381
          :                |||||   ::::::   :::   |||||||
     346 s.................AspCysLeuArgGlyAlaSerLeuCysIleM 357

    1382 TGGAGGAAACAGGGGTGTCCCATTCTCGA......AAATTTTCAAAGTGC 1425
          ||   :::   |||::::::   :::|||      :::|||||||
     357 et...ArgProGlyLeuThrLysGlyArgSerTyrGluPheSerAspAsp 372

    1426 AGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCT 1475
          |||:::      |||:::   ||||||::::::            |||::
     373 SerMetHisTyrTyrGluArgPheLeuLysGlnTyrLysProGlnCysIl 389

    1476 TTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAA.........TGTG 1516
          ::::|||:::|||   ::::::   :::|||::::::          |
     389 eLeuAsnLysProLeuArgIle...AspProValSerThrProValSerG 405

    1517 GAAATGGATACGTGGAAGCTGGGGAGGAG 1545
          |||||      :::|||||||||||||||
     405 lyAsnGluLeuLeuGluAlaGlyGluGlu 414

seq_name: SwissProt_39:ATS7_HUMAN

seq_documentation_block:
ID   ATS7_HUMAN     STANDARD;      PRT;   997 AA.
```

```
AC   Q9UKP4;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ADAM-TS 7 PRECURSOR (EC 3.4.24.-) (A DISINTEGRIN AND METALLOPROTEINASE
DE   WITH THROMBOSPONDIN MOTIFS 7) (ADAMTS-7) (ADAM-TS7).
GN   ADAMTS7.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   MEDLINE=99395124; PubMed=10464288;
RA   [redacted]Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
RT   "ADAM-TS5, ADAM-TS6, and ADAM-TS7, Novel Members of a New Family of
RT   Zinc Metalloproteases.";
RL   J. Biol. Chem. 274:25555-25563(1999).
CC   -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: EXPRESED IN HEART, BRAIN, PLACENTA, LUNG,
CC       LIVER, SKELETAL MUSCLE, KIDNEY, PANCREAS.
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF140675; AAD56358.1; -.
DR   HSSP; P15167; 1ATL.
DR   MIM; 605009; -.
DR   InterPro; IPR000130; -.
     InterPro; IPR000884; -.
     InterPro; IPR001590; -.
     InterPro; IPR002870; -.
     Pfam; PF00090; tsp_1; 1.
     Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50092; TSP1; 1.
DR   PROSITE; PS00427; DISINTEGRINS; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Repeat; Extracellular matrix.
FT   SIGNAL       1    ?      POTENTIAL.
FT   PROPEP       ?    232    BY SIMILARITY.
FT   CHAIN      233    997    ADAM-TS 7.
FT   SITE       204    204    CYSTEINE SWITCH (POTENTIAL).
FT   METAL      388    388    ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE   389    389    BY SIMILARITY.
FT   METAL      392    392    ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL      398    398    ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN     462    537    DISINTEGRIN-LIKE.
FT   DOMAIN     538    594    TSP-TYPE 1 1.
FT   DOMAIN     595    697    CYS-RICH.
FT   DOMAIN     698    914    SPACER.
FT   DOMAIN     915    990    TSP-TYPE 1 2.
FT   CARBOHYD    94     94    N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   693    693    N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   778    778    N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   997 AA;  109694 MW;  6587044ED02FC104 CRC64;
```

```
alignment_scores:
            Quality:  324.00              Length:    860
              Ratio:  0.850                 Gaps:     43
 Percent Similarity:  44.302   Percent Identity: 21.628

alignment_block:
 US-09-634-252A-2 x ATS7_HUMAN   ..

 Align seg 1/1  to: ATS7_HUMAN  from: 1  to: 997

     85 CCCGCCGGCTCGGTGCCTGCCAGC...GCCCCGGCCCGCACGCCGCCCTG 131
        |||:::|||      |||   |||   ||||||         |||
      2 ProGlyGlyProSerProArgSerProAlaProLeuLeuArgPro..... 16

    132 CCGCCTGCTTCTCGTCCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTCCC 181
                |||:::|||||||||| |||      :::|||      |||
     17 ..........LeuLeuLeuLeuLeu.CysAlaLeuAlaProGlyAlaPro 29

    182 GGCCCCGCGCCTGGGGGGCTGCTGCGCCCAGCGCTCCGCATTGGAATGAA 231
        ||||||||||||||||
     30 GlyProAlaProGly..................................... 34

    232 ACTGCAGAAAAAAATTTGGGAGTCCTGGCAGATGAAGACAATACATTGCA 281

     34 ..................................................... 34

    282 ACAGAATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAA 331

     34 ..................................................... 34

    332 TCACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGAAAGC 381
                                        :::|||:::   |||  ||
     35 ..........................ArgAlaThrGluGlyArgAlaAl 42

    382 CCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAA 431
        ||||                    |||   |||::::::   :::::::::
     42 aLeuAspIleValHisProValArgValAspAlaGlyGlySerPheLeuS 59

    432 GGCTGTCCATCTGGCCCAGGCAAGC.......................... 456
             ::::::|||||||||
     59 erTyrGluLeuTrpProArgAlaLeuArgLysArgAspValSerValArg 75

    457 ..................TTCCAGATTGAAGCCTTCGGCTCCAAATTCAT 488
                           ::::::::::::      |||   ::::::
     76 ArgAspAlaProAlaPheTyrGluLeuGlnTyrArgGlyArgGluLeuAr 92

    489 TCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGA 538
         ::::::|||      |||:::   |||||||:::      :::|||:::
     92 gPheAsnLeuThrAlaAsnGlnHisLeuLeuAlaProGlyPheValSerG 109

    539 TTCACTACGAAAATGGGAAACCACAGTACTCTAAGGGTGGAGAGCAC... 585
              :::::::|||             :::   |||      |||
    109 luThrArgArgArgGly.............GlyLeuGlyArgAlaHisIle 121

    586 ..................TGTTACTACCATGGAAGCATCAGAGGC..... 612
                           |||:::      |||:::::::::
    122 ArgAlaHisThrProAlaCysHisLeuLeuGlyGluValGlnAspProGl 138

    613 .GTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCA 661
         ::::::   :::   :::|||:::|||:::|||:::|||||||   |||:
    138 uLeuGluGlyGlyLeuAlaAlaIleSerAlaCysAspGlyLeuLysGlyV 155

    662 TGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTT 711
        ::|||:::   ::::::      |||:::|||||||||||||:::  .:::
    155 alPheGlnLeuSerAsnGluAspTyrPheIleGluProLeuAspSerAla 171

    712 CATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAACCTTGGC 761
            :::   :::   ::::::||||||::::::
    172 ProAlaArgProGlyHisAlaGlnProHisValVal.............. 183

    762 AGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAGGTGACC 811
```

```
          |||:::::::|||      :::         :::||||||||||:
  184 .......TyrLysArgGlnAlaProGluArgLeuAlaGlnArgGlyAspS 198

  812 AG..............................TGGCCCTTTCTC...... 825
      ::                                :::|||   |||
  198 erSerAlaProSerThrCysGlyValGlnValTyrProGluLeuGluSer 214

  826 ..................TCTGAATTACAGTGGTTGAAAAGAAGGAAGAG 857
                       ::::::   ||||||   :::   |||   ||
  215 ArgArgGluArgTrpGluGlnArgGlnGlnTrpArgArgProArgLeuAr 231

  858 AGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATTTGGAAC 907
      |   ::::::   :::|||:::::::   ::::|||   |||:::::::|||
  231 gArgLeuHis...GlnArgSerValSerLysGlu...LysTrpValGluT 246

  908 TTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCAT 957
        ::::::|||   |||   |||      :::   |||      ::::::
  246 hrLeuValValAlaAspAlaLysMetValGluTyrHis......GlyGln 260

  958 GCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTAT 1007
          ::::::::::::::::::   ::::::::::|||:::|||   :::::
  261 ProGlnValGluSerTyrValLeuThrIleMetAsnMetValAlaGlyLe 277

 1008 TTACAAGGAGCAGCTC.................AACACCAGGGTTGTCC 1039
      ::::   :::                       |||   |||   |
  277 uPheHisAspProSerIleGlyAsnProIleHisIleThrIleValArgL 294

 1040 TGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACC 1089
      |||||   ::|:|||         |||:::::::::::   ||||||
  294 euValLeuLeuGlu......AspGluGluGluAspLeuLysIleThrHis 308

 1090 AACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATTAA 1139
      :::       :::   |||   :::|||   |||:::::::::   |||::
  309 HisAlaAspAsnThrLeuLysSerPheCysLysTrpGlnLysSerIleAs 325

 1140 GCAGCATGCTGATGCT......GTGCAC..................... 1161
      ::::   :::||||||      :::|||
  325 nMetLysGlyAspAlaHisProLeuHisHisAspThrAlaIleLeuLeuT 342

 1162 ...........CTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGT 1200
                 |||   :::   ::::::         :::      :::
  342 hrArgLysAspLeuCysAlaAlaMetAsnArgProCysGluThrLeuGly 358

 1201 CTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAA 1250
      ||||||:::   :::|||:::|||:::      |||:::  ::::::||
  359 LeuSerHisValAlaGlyMetCysGlnProHisArgSerCysSerIleAs 375

 1251 TGAG...TATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCC 1297
      ||||      |||||||||:::|||          :::::::::::::|
  375 nGluAspThrGlyLeuProLeuAlaPhe......ThrValAlaHisGluL 390

 1298 TGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAA 1347
      ||:::::::::::::|||||||||   :::   |||::::::
  390 euGlyHisSerPheGlyIleGlnHisAspGlySerGlyAsn......... 403

 1348 TGTGACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGT 1397
         ||||||                          |||   :::|||
  404 ...AspCys.........................GluProValGly.. 409

 1398 GTCCCATTCTCGAAAATTT.............................. 1416
             :::|||   |||
  410 .......LysArgProPheIleMetSerProGlnLeuLeuTyrAspAlaA 424

 1417 ..............TCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTA 1452
                    |||:::||||||                  ||||||
  424 laProLeuThrTrpSerArgCysSerArgGlnTyrIleThrArgPheLeu 440

 1453 CAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAG......CT 1496
      :::||||||   |||   ||||||   :::   |||:::|||      ::
  441 AspArgGlyTrpGlyLeuCysLeuAspAspProProAlaLysAspIleIl 457

 1497 ATTTGAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGT 1546
      :      |||:::                  |||      |||
```

```
  457 eAspPheProSer.................ValProProGlyValLeuT 468

 1547 GTGATTGTGGTTTTCATGTGGAATGCTATGGATTATGCTGTAAGAAATGT 1596
        |||   :::                                 :::|||
  468 yrAspValSer.............................HisGlnCys 474

 1597 TCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATAC 1646
         |||:::   ||||||:::   :::
  475 ArgLeuGlnTyrGlyAlaTyrSerAla..................... 483

 1647 CTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACG 1696
                                        |||:::|||      |||
  484 ........................PheCysGluAspMetAspAsnV 491

 1697 AGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCA 1746
        |||      |||   :::|||:::      :::      |||
  491 alCysHis...ThrLeuTrpCysSerValGlyThr...ThrCysHisSer 505

 1747 AATCTTCATAAGCAA...GACGGATATGCATGCAATCAAAATCAGGGCCG 1793
      :::|||            ||||||      |||:::::|||:::
  506 LysLeuAspAlaAlaValAspGlyThrArgCysGlyGluAsnLysTrp.. 521

 1794 CTGCTACAATGGCGAGTGCAAGACC.........AGAGACAACCAGTGTC 1834
        |||   :::|||||||||               |||   :::      :
  522 .CysLeuSerGlyGluCysValProValGlyPheArgProGluAlaValA 538

 1835 AGTACATCTGGGGA.......................ACAAAGGCTGCA 1860
      ::      |||:::                       ::::::::::::
  538 spGlyGlyTrpSerGlyTrpSerAlaTrpSerIleCysSerArgSerCys 554

 1861 GGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAA 1910
      |||                  |||:::      |||:::   |||   ||
  555 GlyMetGlyValGlnSerAlaGluArgGlnCysThrGlnProThrProLy 571

 1911 GGGAAAC.......................................... 1917
      |   :::
  571 sTyrLysGlyArgTyrCysValGlyGluArgLysArgPheArgLeuCysA 588

 1918 .................TGCGGGAAGGATGGAGACCGGTGGATTCAGTGC 1950
                       :::|||:::   :::   |||   :::||||||
  588 snLeuGlnAlaCysProAlaGlyArgProSerPheArgHisValGlnCys 604

 1951 AGCAAACATGAT......GTGTTCTGTGGATTCTTACTCTGTACCAATCT 1994
      |||      |||      ::::::   |||   |||
  605 SerHisPheAspAlaMetLeuTyrLysGlyGlnLeu............Hi 617

 1995 TACTCGAGCTCCACGTATTGGTCAACTT.........CAGGGTGAGATCA 2035
        |||   :::|||   ::::::::::::         :::   :::
  617 sThrTrpValProValValAsnAspValAsnProCysGluLeuHisCysA 634

 2036 TTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCC 2085
        |||::::::   |||         ::::::   |||:::
  634 rgProAlaAsnGluTyrPheAlaLysLysLeuArgAspAla......... 647

 2086 CATGTAGTTTTAGATGATGATACGGATGTGGGCTATGTAGAAGATGGAAC 2135
                                        |||   ||||||||
  648 ................................ValValAspGlyTh 652

 2136 GCCATGT............GGCCCGTCTATGATGTGTTTAGATCGGAAGT 2173
      |||||||           :::         :::|||::::::      |
  652 rProCysTyrGlnValArgAlaSerArgAspLeuCysIleAsnGlyIleC 669

 2174 GCCTACAAATTCAAGCC...CTAAATATGAGCAGCTGTCCACTCGATTCC 2220
      ||  ::::::   :::   ::::::::::::|||      :::::::::
  669 ysLysAsnValGlyCysAspPheGluIleAspSerGlyAlaMetGluAsp 685

 2221 AAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTG 2270
      :::      ||||||   |||:::|||               :::||||||
  686 ArgCysGlyValCysHisGlyAsnGly..............SerThrCy 697

 2271 CATTTGTGATTTCACCTGGGCAGGGACA 2298
      |           |||   :::||||||
  697 s.........HisThrValSerGlyThr 703
```

```
seq_name: SwissProt_39:ATS1_HUMAN

seq_documentation_block:
ID   ATS1_HUMAN      STANDARD;      PRT;   967 AA.
AC   Q9UHI8; Q9UP80; Q9UH83;
DT   30-MAY-2000 (Rel. 39, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ADAM-TS 1 PRECURSOR (EC 3.4.24.-) (A DISINTEGRIN AND METALLOPROTEINASE
DE   WITH THROMBOSPONDIN MOTIFS 1) (ADAMTS-1) (ADAM-TS1) (METH-1).
GN   ADAMTS1 OR METH1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Casas C., Pritchard M.A., Estivill X., Arbones M.L.;
RT   "Cloning, characterization and mapping on human chromosome 21 of the
RT   orthologue of murine Adamts-1.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A., AND FUNCTION.
RC   TISSUE=Heart;
RX   MEDLINE=99367466; PubMed=10438512;
RA   Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,
RA   Lombardo M., Iruela-Arispe M.L.;
RT   "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new
RT   family of proteins with angio-inhibitory activity.";
RL   J. Biol. Chem. 274:23349-23357(1999).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Endothelial cells;
RX   MEDLINE=20247184; PubMed=10785405;
RA   Glienke J., Schmitt A.O., Pilarsky C., Hinzmann B., Weiss B.,
RA   Rosenthal A., Thierauch K.H.;
RT   "Differential gene expression by endothelial cells in distinct
RT   angiogenic states.";
RL   Eur. J. Biochem. 267:2820-2830(2000).
RN   [4]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20289799; PubMed=10830953;
RA   Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA   Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA   Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA   Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA   Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA   Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA   Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA   Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA   Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA   Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA   Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA   Lehrach H., Reinhardt R., Yaspo M.-L.;
RT   "The DNA sequence of human chromosome 21.";
RL   Nature 405:311-319(2000).
CC   -!- FUNCTION: CLEAVES AGGRECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC       INVOLVED IN ITS TURNOVER (BY SIMILARITY). HAS ANGIOGENIC INHIBITOR
CC       ACTIVITY. ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH
CC       VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC       CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE.
CC   -!- CATALYTIC ACTIVITY: CLEAVES AGGRECAN AT THE 1933-GLU-|-LEU-1939
CC       SITE, WITHIN THE CHONDROITIN SULFATE ATTACHEMENT DOMAIN.
CC   -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF170084; AAF15317.1; -.
DR   EMBL; AF060152; AAD48080.1; ALT_INIT.
DR   EMBL; AF207664; AAF23772.1; -.
DR   EMBL; AP001697; BAA95502.1; -.
DR   MIM; 605174; -.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR000884; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF00090; tsp_1; 3.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR   PROSITE; PS50092; TSP1; 2.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Repeat; Extracellular matrix; Heparin-binding.
FT   SIGNAL        1     49       POTENTIAL.
FT   PROPEP       50    252       BY SIMILARITY.
FT   CHAIN       253    967       ADAM-TS 1.
FT   SITE        198    198       CYSTEINE SWITCH (POTENTIAL).
FT   METAL       401    401       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    402    402       BY SIMILARITY.
FT   METAL       405    405       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       411    411       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN      476    559       DISINTEGRIN-LIKE.
FT   DOMAIN      560    616       TSP-TYPE 1 1.
FT   DOMAIN      617    724       CYS-RICH.
FT   DOMAIN      725    849       SPACER.
FT   DOMAIN      850    908       TSP-TYPE 1 2.
FT   DOMAIN      909    967       TSP-TYPE 1 3.
FT   DOMAIN      843    846       POLY-LYS.
FT   CARBOHYD    547    547       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    720    720       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    764    764       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    227    227       P -> A (IN REF. 3 AND 4).
FT   CONFLICT    468    468       Q -> H (IN REF. 1).
FT   CONFLICT    561    561       S -> N (IN REF. 1).
SQ   SEQUENCE   967 AA;  105383 MW;  C189389324741ED1 CRC64;


alignment_scores:
            Quality:  309.50              Length:    972
              Ratio:  0.742                 Gaps:     47
 Percent Similarity:  42.901   Percent Identity: 19.856

alignment_block:
 US-09-634-252A-2 x ATS1_HUMAN   ..

 Align seg 1/1  to: ATS1_HUMAN  from: 1  to: 967

      70 GGCCCCCAACGCGGCCCCGCCGGCTCGGTGCCTGCCAGCGCCCCGGCCCG 119
         |||  :::   |||   |||      :::|||:::|||
      15 GlySerAspMetGlyAsnAlaGluArgAlaProGlySerArgSerPheGl 31

     120 CACGCCGCCCTGCCGCCTGCTTCTCGTCCTTCTCCTGCTGCCTCCGCTCG 169
               |||      |||||||||:::      ||||||
      31 yProValProThrLeuLeuLeuLeuAlaAlaAlaLeuLeuAlaValSerA 48

     170 CCGCCTCGTCCCGGCCCCGCGCCTGGGGGGCTGCTGCGCCCAGCGCTCCG 219
           |||   :::||||||
```

```
 48 spAlaLeuGlyArgPro..............................  53

220 CATTGGAATGAAACTGCAGAAAAAAATTTGGGAGTCCTGGCAGATGAAGA 269
                                          ::::::|||||
 54 ......................................SerGluGluAs  57

270 CAATACATTGCAACAGAATAGCAGCAGTAATATCAGTTACAGCAATGCAA 319
    |
 57 p..............................................  57

320 TGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACATCAACCAA 369
       :::|||::::::::::|||                      :::
 58 .....GluGluLeuValValPro.......................Glu  64

370 GACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCA 419
     ::::::::::|||   |||
 65 LeuGluArgAlaProGlyHis...........................  71

420 AAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAG 469
                              :::   :::   ::::::::::::|
 72 .......................GlyThrThrArgLeuArgLeuHisA  80

470 CCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTG 519
    |||||   :::::::::   |||:::|||      ::::::::::::|||
 80 laPheAspGlnGlnLeuAspLeuGluLeuArgProAspSerSerPheLeu  96

520 TCTTCTGATTATGTG......GAGATTCACTACGAAAATGGGAAA..... 558
    :::      ::::::     ::::::     ::::::|||:::.
 97 AlaProGlyPheThrLeuGlnAsnValGlyArgLysSerGlySerGluTh 113

559 .CCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCA 607
     |||      :::          ||||||:::|||   |||:::::::
113 rProLeuProGluThrAspLeuAlaHisCysPheTyrSerGlyThrValA 130

608 GAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCAT 657
    ::|||      :::|||   :::|||||||||   |||:::|||::::::
130 snGlyAspProSerSerAlaAlaAlaLeuSerLeuCysGluGlyValArg 146

658 GGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCT 707
    |||  |||              :::|||:::|||:::||||||
147 GlyAlaPheTyrLeuLeuGlyGluAlaTyrPheIleGlnProLeuProAl 163

708 GGTTCATGATGAGAAAAGCACAGGTCGACCA.................. 738
     :::   ::::::   :::|||:::   |||
163 aAlaSerGluArgLeuAlaThrAlaAlaProGlyGluLysProProAlaP 180

739 ..........CATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAG 777
           |||:::::::::::::::      |||:::   :::
180 roLeuGlnPheHisLeuLeuArgArgAsnArgGlnGlyAspValGlyGly 196

778 CAAATGAAGAATCTCACTATGGAA......................... 801
                :::      |||
197 ThrCysGlyValValAspAspGluProArgProThrGlyLysAlaGluTh 213

802 ................................AGAGGTGACCAGTGG. 816
                                   :::|||   ||||||
213 rGluAspGluAspGluGlyThrGluGlyGluAspGluGlyProGlnTrpS 230

817 ..CCCTTTCTCTCTGAATTACAGTGGTTGAAA.................. 846
      |||           ||||||   :::
230 erProGlnAspProAlaLeuGlnGlyValGlyGlnProThrGlyThrGly 246

847 ......AGAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGA 890
          |||:::||||||   |||:::      |||
247 SerIleArgLysLysArgPheValSerSerHisArg.............. 258

891 AATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGA 940
        |||:::|||   |||:::|||   |||::::::      :::
259 .......TyrValGluThrMetLeuValAlaAspGlnSerMetAlaGluP 273

941 AGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTC 990
      |||   |||:::      |||         :::      ::::::
273 heHisGlySerGlyLeuLysHis.........TyrLeuLeuThrLeuPhe 286

991 AACCTTGTGGATTCTATTTACAAGGAG...CAGCTCAACACCAGGGTTGT 1037
    :::::::::      :::||||||:::   ::::::::::::   |||
287 SerValAlaAlaArgLeuTyrLysHisProSerIleArgAsnSerValSe 303

1038 CCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAG.........ATTG 1078
      ||||||:::|||:::      :::      |||:::          :
303 rLeuValValValLysIleLeuValIleHisAspGluGlnLysGlyProG 320

1079 ACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGG 1128
     :::::|||:::|||   :::      |||:::::::||   :::::::::
320 luValThrSerAsnAlaAlaLeuThrLeuArgAsnPheCysAsnTrpGln 336

1129 CAGCGCATTAAGCAGCAT..............GCTGATGCTGTGCACCT 1163
             |||||||||                 ||||||   |||
337 .........LysGlnHisAsnProProSerAspArgAspAlaGluHisTy 350

1164 CATCTCGCGGGTGACATTTCACTATAAG..................... 1191
        :::   :::   |||      :::
350 rAspThrAlaIleLeuPheThrArgGlnAspLeuCysGlySerGlnThrC 367

1192 .....AGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGA 1236
                :::::::::      |||   ||||||:::   :::|||
367 ysAspThrLeuGlyMetAlaAspValGlyThrValCysAspProSerArg 383

1237 GGAGTTGGTGTG...AATGAGTATGGTCTTCCAATGGCAGTGGCACAAGT 1283
     :::   :::|||   ::::::   ||||||      |||
384 SerCysSerValIleGluAspAspGlyLeuGlnAlaAlaPhe........ 397

1284 ATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAA...CCTT 1330
           :::   |||::::::||||||      ::::::   |||
398 ..........ThrThrAlaHisGluLeuGlyHisValPheAsnMetProH 411

1331 CTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATC 1380
     :::          :::|||            |||   :::|||
411 isAspAspAlaLysGlnCys.........AlaSerLeuAsnGlyValAsn 424

1381 ATGGAGGAA.....................ACAGGGGTGTCCCATTCTCG 1409
     :::::::::                     ::::::::::::||||||::
425 GlnAspSerHisMetMetAlaSerMetLeuSerAsnLeuAspHisSerGl 441

1410 AAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAG 1459
     :   :::|||   ||||||              :::||||||::::::|
441 nProTrpSerProCysSerAlaTyrMetIleThrSerPheLeuAspAsnG 458

1460 GAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAG...CCC 1506
     ||   |||   ||||||:::::::::|||   :::   ::::::   |||
458 lyHisGlyGluCysLeuMetAspLysProGlnAsnProIleGlnLeuPro 474

1507 ACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGG 1556
        :::      :::         :::|||::::::::::|||:::
475 GlyAspLeuProGlyThrSerTyrAspAlaAsnArgGlnCysGlnPheTh 491

1557 TTTTCATGTGGAATGCTATGGATTATGCTGTAAGAAATGTTCCCTCTCCA 1606
      |||                                       ::::
491 rPheGly.....................................GluA 495

1607 ACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTT 1656
     :::::   ||||||
495 spSerLysHisCysPro............................... 500

1657 TTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATAT 1706
                                 ||||||::::::   |||:::
501 ........................AspAlaAlaSerThrCysSer.. 507

1707 TACTGAATATTGTACTGGAGACTCTGGT.........CAGTGC...CCAC 1744
      |||   :::|||||||||   ||||||          |||
508 .ThrLeuTrpCysThrGlyThrSerGlyGlyValLeuValCysGlnThrL 524.

1745 CAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGC 1794
       ::::::         ||||||   :::|||::::::::::::
524 ysHisPheProTrpAlaAspGlyThrSerCysGlyGluGlyLysTrp... 539
```

```
1795 TGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTAC..... 1839
     |||   ||||||:::|||   ::::::   ::::::   ::::::
 540 CysIleAsnGlyLysCysValAsnLysThrAspArgLysHisPheAspTh 556

1840 .....................ATCTGGGGA..............ACAA 1852
                          :::||||||                ::::
 556 rProPheHisGlySerTrpGlyMetTrpGlyProTrpGlyAspCysSerA 573

1853 AGGCTGCAGGG.............................'..... 1863
     ::::::::|||
 573 rgThrCysGlyGlyGlyValGlnTyrThrMetArgGluCysAspAsnPro 589

1864 ............TCTGACAAGTTCTGC.................TATGA 1883
                 :::   |||:::|||                  |||::
 590 ValProLysAsnGlyGlyLysTyrCysGluGlyLysArgValArgTyrAr 606

 884 AAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGCAG 1933
     ::::   |||   |||            :::|||   ||||||
 606 gSerCysAsnLeuGluAspCysProAspAsnAsn...GlyLysThr...P 621

1934 ACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTC 1983
       |||      ||||||:::   |||:::   |||           :::
 621 heArgGluGluGlnCysGluAlaHisAsnGluPheSerLysAlaSerPhe 637

1983 ..................................................  1983

 638 GlySerGlyProAlaValGluTrpIleProLysTyrAlaGlyValSerPr 654

1984 ..........TGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACTTC 2023
               |||      :::   :::|||      ||||||   :::
 654 oLysAspArgCysLysLeuIleCysGlnAlaLysGlyIleGlyTyrPhe. 670

2024 AGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTGAC 2073
                        |||      |||   :::|||:::
 671 .....................PheValLeuGlnProLysValVal.... 678

2074 TGCAGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTATGT 2123

 678 ..................................................  678

2124 AGAAGATGGAACGCCATGTGGCCCG......TCTATGATGTGTTTAGATC 2167
         ||||||||||||||||:::|||      :::   :::|||::::::
 679 ....AspGlyThrProCysSerProAspSerThrSerValCysValGlnG 694

2168 GGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCA...... 2211
     :::|||::::::                  :::::::|||
 694 lyGlnCysValLys.................AlaGlyCysAspArgIle 704

2212 CTCGATTCCAAGGGTAAA.............GTCTGTTCGGGCCATGG 2246
     :::||||||||||   |||             ||||||:::|||:::||
 705 IleAspSerLysLysLysPheAspLysCysGlyValCysGlyGlyAsnGl 721

2247 GGTGTGTAGTAATGAAGCCACCTGC..................... 2271
     |               :::||||||
 721 y.............SerThrCysLysLysIleSerGlySerValThrS 733

2272 ......................ATTTGTGATTTCACCTGGGCAGGGACA 2298
                          |||      :::      :::::::|||
 733 erAlaLysProGlyTyrHisAspIleIleThrIleProThrGlyAlaThr 749

2299 GATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAAGGATGA 2348
     :::   ::::::::::::      ||||||   :::      ::::::::
 750 AsnIleGluValLysGln......ArgAsnGlnArgGlySerArgAsnAs 764

2349 AGGACCC............AAGGGTCCTAGTGCCACCAATCTCATAATAG 2386
     :|||              |||:::   :::::::|||   ::::::   |
 764 nGlySerPheLeuAlaIleLysAlaAlaAspGlyThrTyrIleLeuAsnG 781

2387 GC..............TCCATCGCTGGTGCCATCCTGGTAGCAGCTATT 2421
     ||              ::::::        |||:::      ::::::
 781 lyAspTyrThrLeuSerThrLeuGluGlnAspIleMetTyrLysGlyVal 797

2422 GTCCTTGGGGGCACAGGCTGGGGATTTAAAAATGTCAAGAAGAGAAG3TT 2471
```

```
         ||||||      :::|||   :::            :::   |||   ||
  798 ValLeuArgTyrSerGlySerSerAlaAlaLeuGluArgIleArgSerPh 814

 2472 CGATCCTACTCAGCAA 2487
      |:::|||   ::::::
  814 eSerProLeuLysGlu 819
```

seq_name: SwissProt_39:ATS1_MOUSE

seq_documentation_block:

```
ID   ATS1_MOUSE     STANDARD;      PRT;   968 AA.
AC   P97857; O54768;
DT   30-MAY-2000 (Rel. 39, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ADAM-TS 1 PRECURSOR (EC 3.4.24.-) (A DISINTEGRIN AND METALLOPROTEINASE
DE   WITH THROMBOSPONDIN MOTIFS 1) (ADAMTS-1) (ADAM-TS1).
GN   ADAMTS1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=129/SVJ;
RX   MEDLINE=98110583; PubMed=9441751;
RA   Kuno K., Lizasa H., Ohno S., Matsushima K.;
RT   "The exon/intron organization and chromosomal mapping of the mouse
RT   ADAMTS-1 gene encoding an ADAM family protein with TSP motifs.";
RL   Genomics 46:466-471(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97150761; PubMed=8995297;
RA   Kuno K., Kanada N., Nakashima E., Fujiki F., Ichimura F.,
RA   Matsushima K.;
RT   "Molecular cloning of a gene encoding a new type of metalloproteinase-
RT   disintegrin family protein with thrombospondin motifs as an
RT   inflammation associated gene.";
RL   J. Biol. Chem. 272:556-562(1997).
RN   [3]
RP   CHARACTERIZATION, AND MUTAGENESIS OF GLU-403.
RX   MEDLINE=99303657; PubMed=10373500;
RA   Kuno K., Terashima Y., Matsushima K.;
RT   "ADAMTS-1 is an active metalloproteinase associated with the
RT   extracellular matrix.";
RL   J. Biol. Chem. 274:18821-18826(1999).
RN   [4]
RP   FUNCTION.
RX   MEDLINE=20389568; PubMed=10930576;
RA   Kuno K., Okada Y., Kawashima H., Nakamura H., Miyasaka M.,
RA   Ohno H., Matsushima K.;
RT   "ADAMTS-1 cleaves a cartilage proteoglycan, aggrecan.";
RL   FEBS Lett. 478:241-245(2000).
RN   [5]
RP   FUNCTION, AND INDUCTION.
RX   MEDLINE=20243757; PubMed=10781075;
RA   Robker R.L., Russell D.L., Espey L.L., Lydon J.P., O'Malley B.W.,
RA   Richards J.S.;
RT   "Progesterone-regulated genes in the ovulation process: ADAMTS-1 and
RT   cathepsin L proteases.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:4689-4694(2000).
CC   -!- FUNCTION: CLEAVES AGGRECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC       INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY
CC       SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH
CC       VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC       CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
CC       SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: CLEAVES AGGRECAN AT THE 1691-GLU-|-LEU-1692
CC       SITE, WITHIN THE CHONDROITIN SULFATE ATTACHEMENT DOMAIN.
CC   -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX.
CC   -!- INDUCTION: INDUCED IN VITRO IN COLON ADENOCARCINOMA CELLS BY
```

```
CC       INTERLEUKIN-1, OR IN VIVO IN KIDNEY AND HEART BY
CC       LIPOPOLYSACCHARIDE. ALSO INDUCED BY LH STIMULATION IN GRANULOSA
CC       CELLS OF PREOVULATORY FOLLICLES.
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB001735; BAA24501.1; ALT_INIT.
DR   EMBL; D67076; BAA11088.1; ALT_FRAME.
DR   MEROPS; M12.222; -.
DR   MGD; MGI:109249; Adamts1.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR000884; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF00090; tsp_1; 3.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR   PROSITE; PS50092; TSP1; 3.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Repeat; Extracellular matrix; Heparin-binding.
FT   SIGNAL       1     48       POTENTIAL.
FT   PROPEP      49    253
FT   CHAIN      254    968       ADAM-TS 1.
FT   SITE       206    206       CYSTEINE SWITCH (POTENTIAL).
FT   METAL      402    402       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE   403    403
FT   METAL      406    406       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL      412    412       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN     477    559       DISINTEGRIN-LIKE.
FT   DOMAIN     560    617       TSP-TYPE 1 1.
FT   DOMAIN     618    725       CYS-RICH.
FT   DOMAIN     726    850       SPACER.
FT   DOMAIN     851    909       TSP-TYPE 1 2.
FT   DOMAIN     910    968       TSP-TYPE 1 3.
FT   DOMAIN     195    199       POLY-ARG.
FT   CARBOHYD   548    548       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   721    721       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   765    765       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   783    783       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   946    946       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   MUTAGEN    403    403       E->Q: LOSS OF ACTIVITY.
FT   CONFLICT   335    335       N -> S (IN REF. 2).
FT   CONFLICT   425    425       T -> S (IN REF. 2).
SQ   SEQUENCE   968 AA;  105841 MW;  42EBDA55499FB6C1 CRC64;


alignment_scores:
            Quality:  304.50              Length:   1001
              Ratio:  0.725                Gaps:     49
 Percent Similarity:  41.958   Percent Identity: 20.180

alignment_block:
 US-09-634-252A-2 x ATS1_MOUSE   ..

 Align seg 1/1  to: ATS1_MOUSE  from: 1  to: 968

       4 AAGCCGCCCGGCAGCAGCTCGCGGCAGCCGCCCCTGGCGGGCTGCAGCCT 53
```

```
   |||   |||   :::|||   :::|||   |||         ||||||
 4 LysValProLeuGlySerArgLysGlnLysPro.....CysSer.. 16

54 TGCCGGCGCTTCCTGCGGCCCCCAACGCGGCCCCGCCGGCTCGGTGCCTG 103
                                          |||:::|||
17 ..............................AspMetGlyAspValGlnA 23

104 CCAGCGCCCCGGCCCGC...ACGCCGCCCTGCCGCCTGCTTCTCGTCCTT 150
     :::|||   :::|||   :::     :::::::::||||||:::|||
 23 rgAlaAlaArgSerArgGlySerLeuSerAlaHisMetLeuLeuLeuLeu 39

151 CTC..............CTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCC 185
    |||              ||||||        :::|||   :::|||||
 40 LeuAlaSerIleThrMetLeuLeuCysAlaArgGlyAlaHisGlyArgPr 56

186 CCGCGCCTGGGGGGCTGCTGCGCCCAGCGCTCCGCATTGGAATGAAACTG 235
    |
 56 o..................................................... 56

236 CAGAAAAAAATTTGGGAGTCCTGGCAGATGAAGACAATACATTGCAACAG 285
                          ::::::||||||
 57 ......................ThrGluGluAsp.............. 60

286 AATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAATCAC 335
                                                :::|||:::::
 61 ..........................................GluGluLeuVa 64

336 ACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGAAAGCCCTT 385
    :|||||||||                              :::::::::|||
 64 lLeuProSerLeu.......................GluArgAlaProG 73

386 ATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAGGCT 435
      |||         |||
 73 lyHisAspSerThrThr..................................... 78

436 GTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATT 485
                  :::   :::::::::::|||||||||:::::::
 79 ..............ThrArgLeuArgLeuAspAlaPheGlyGlnGlnLe 90

486 CATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGG 535
    :   |||   |||      ::::::|||:::|||:::
 90 uHisLeuLysLeuGlnProAspSerGlyPheLeuAla.............. 102

536 AGATTCACTACGAAAATGGGAAACCACAGTAC..................... 567
                           |||   :::
103 ......................ProGlyPheThrLeuGlnThrValGly 111

568 ...TCTAAGGGTGGAGAG.................................. 582
       |||   |||:::|||
112 ArgSerProGlySerGluAlaGlnHisLeuAspProThrGlyAspLeuAl 128

583 .CACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGG 631
     ||||||:::|||   |||:::::::::|||         |||   :::|
128 aHisCysPheTyrSerGlyThrValAsnGlyAspProGlySerAlaAlaA 145

632 CTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATGATACCTTC 681
    |||||||   |||:::|||:::::|||   |||      :::
145 laLeuSerLeuCysGluGlyValArgGlyAlaPheTyrLeuGlnGlyGlu 161

682 GTGTATATGATAGAGCCA.........................CTAGA 704
       ::::::|||:::|||
162 GluPhePheIleGlnProAlaProGlyValAlaThrGluArgLeuAlaPr 178

705 GCTGGTTCATGATGAGAAAAGCACAGGTCGACCA......CATATAATCC 748
       |||   :::|||:::|||:::::||||||      ||||||::::
178 oAlaValProGluGluGluSerSerAlaArgProGlnPheHisIleLeuA 195

749 AGAAAACCTTGGCAGGACAGTATTCTAAGCAA.........ATGAAGAAT 789
    :::::          |||:::   :::   :::         |||   :::
195 rgArgArgArgArgGlySerGlyGlyAlaLysCysGlyValMetAspAsp 211

790 CTCACTATGGAAAGAGGTGAC........................... 810
       |||:::      :::|||
```

```
 212 GluThrLeuProThrSerAspSerArgProGluSerGlnAsnThrArgAs 228

 811 .CAGTGGCCCTTTCTCTCTGAATTA..................... 834
      |||||||||      :::
 228 nGlnTrpProValArgAspProThrProGlnAspAlaGlyLysProSerG 245

 835 ..CAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGTATA 882
                  |||:::||||||   |||:::      |||
 245 lyProGlySerIleArgLysLysArgPheValSerSerProArg..... 259

 883 TTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAAC 932
                   |||:::|||   |||:::|||   |||::::::
 260 ..............TyrValGluThrMetLeuValAlaAspGlnSerMe 271

 933 GTATAAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGT 982
             |||   |||:::      |||           :::      :
 271 tAlaAspPheHisGlySerGlyLeuLysHis.........TyrLeuLeuT 285

 983 CCGTGGTCAACCTTGTGGATTCTATTTACAAGGAG...CAGCTCAACACC 1029
     :::::   :::::::::      :::||||||:::   ::::::::::::
 285 hrLeuPheSerValAlaAlaArgPheTyrLysHisProSerIleArgAsn 301

1030 AGGGTTGTCCTGGTGGCTGTAGAG.........ACCTGGACTGAGAAGGA 1070
        :::   ||||||:::|||:::             :::   |||:::
 302 SerIleSerLeuValValValLysIleLeuValIleTyrGluGluGlnLy 318

1071 TCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAA 1120
           ::::::|||:::|||   :::      |||::::::|||   :
 318 sGlyProGluValThrSerAsnAlaAlaLeuThrLeuArgAsnPheCysA 335

1121 AATACCGGCAG.....................CGCATTAAGCAGCATGCT 1149
     ::::::::::::                   |||      :::|||
 335 snTrpGlnLysGlnHisAsnSerProSerAspArgAspProGluHisTyr 351

1150 GATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGC.. 1197
     |||::::::   |||::::::|||      :::      :::::::::
 352 AspThrAlaIleLeuPheThrArgGlnAspLeuCysGlySerHisThrCy 368

1198 ..........AGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAG 1237
               :::::::::      |||   ||||||:::   :::| |:
 368 sAspThrLeuGlyMetAlaAspValGlyThrValCysAspProSerArgS 385

1238 GAGTTGGTGTG...AATGAGTATGGTCTTCCAATGGCAGTGGCACAAGTA 1284
     ::   :::|||   ::::::   ||||||      |||
 385 erCysSerValIleGluAspAspGlyLeuGlnAlaAlaPhe......... 398

1285 TTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAA...CCTTC 1331
              ::   |||::::::||||||      ::::::   |||
 399 .........ThrThrAlaHisGluLeuGlyHisValPheAsnMetProHi 412

1332 TAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCA 1381
      :::            |||             |||   :::|||
 412 sAspAspAlaLysHisCys.........AlaSerLeuAsnGlyValThrG 426

1382 TGGAGGAA.....................ACAGGGGTGTCCCATTCTCGA 1410
     ::::::                      ::::::::::::||||||:::
 426 lyAspSerHisLeuMetAlaSerMetLeuSerSerLeuAspHisSerGln 442

1411 AAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGG 1460
        :::|||   ||||||                 :::||||||:::::||
 443 ProTrpSerProCysSerAlaTyrMetValThrSerPheLeuAspAsnGl 459

1461 AGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAG...CCCA 1507
     |   |||   ||||||:::::::::|||   :::   ::::::   |||:
 459 yHisGlyGluCysLeuMetAspLysProGlnAsnProIleLysLeuProS 476

1508 CGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGT 1557
     :::::      :::      :::|||:::::::::|||:::
 476 erAspLeuProGlyThrLeuTyrAspAlaAsnArgGlnCysGlnPheThr 492

1558 TTTCATGTGGAATGCTATGGATTATGCTGTAAGAAATGTTCCCTCTCCAA 1607
     |||      |||
 493 PheGlyGluGlu..................................... 496
```

```
1608 CGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTT 1657

 496 ..................................................... 496

1658 TTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATT 1707
                ::   :::|||   ||||||::::::   |||
 497 ...........SerLysHisCysProAspAlaAlaSerThrCys...Thr 508

1708 ACTGAATATTGTACTGGAGACTCTGGT.........CAGTGC...CCACC 1745
     |||   :::|||||||||   ||||||          |||
 509 ThrLeuTrpCysThrGlyThrSerGlyGlyLeuLeuValCysGlnThrLy 525

1746 AAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCT 1795
      ::::::         ||||||   :::|||:::::::::::::      |
 525 sHisPheProTrpAlaAspGlyThrSerCysGlyGluGlyLysTrp...C 541

1796 GCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATC... 1842
     ||   :::|||:::|||   ::::::   ::::::   ::::::
 541 ysValSerGlyLysCysValAsnLysThrAspMetLysHisPheAlaThr 557

1843 ..............TGGGGAACAAAGGCTGCAGGGTCTGAC........ 1869
                    ||||||      :::      :::|||
 558 ProValHisGlySerTrpGlyProTrpGlyProTrpGlyAspCysSerAr 574

1869 ..................................................... 1869

 574 gThrCysGlyGlyGlyValGlnTyrThrMetArgGluCysAspAsnProV 591

1870 ................AAGTTCTGC.................TATGAA 1884
                    |||:::|||                 |||:::
 591 alProLysAsnGlyGlyLysTyrCysGluGlyLysArgValArgTyrArg 607

1885 AAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGA 1934
     :::   |||   |||            :::|||   ||||||
 608 SerCysAsnIleGluAspCysProAspAsnAsn...GlyLysThr...Ph 622

1935 CCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCT 1984
      |||      ||||||:::   |||:::
 622 eArgGluGluGlnCysGluAlaHisAsn..................... 631

1985 GTACCAATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATC 2034
                                                  |||:::
 632 .....................................GluPhe 633

2035 ATTCCAACTTCCTTCTACCATCAAGGCCGGGTG.................. 2067
          :::||||||   ::::::      |||
 634 SerLysAlaSerPheGlyAsnGluProThrValGluTrpThrProLysTy 650

2068 ............................ATTGACTGCAGTGGTGCCC 2086
                                    :::   |||
 650 rAlaGlyValSerProLysAspArgCysLysLeuThrCys......... 663

2087 ATGTAGTTTTAGATGATGATACGGATGTGGGCTAT............... 2121
                  :::          :::||||||
 664 ..............GluAlaLysGlyIleGlyTyrPhePheValLeuGln 675

2122 ......GTAGAAGATGGAACGCCATGTGGCCCG......TCTATGATGTG 2159
           |||   ||||||||||||||||:::|||      :::   :::||
 676 ProLysValValAspGlyThrProCysSerProAspSerThrSerValCy 692

2160 TTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTC 2209
     |::::::   :::|||::::::                  ::::::|||
 692 sValGlnGlyGlnCysValLys..................AlaGlyCysA 703

2210 CA......CTCGATTCCAAGGGTAAA..............GTCTGTTCG 2238
            :::|||||||||   |||               ||||||:::
 703 spArgIleIleAspSerLysLysLysPheAspLysCysGlyValCysGly 719

2239 GGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTG 2288
     |||:::|||         ::::::   :::   |||         |||
 720 GlyAsnGlySerThrCysLysLysMetSerGlyIleValThrSerThrAr 736
```

```
    2289 GGCAGGGACAGATTGCAGTATCCGGGATCCAGTT..... 2322
          |||           :::     |||:::
     736 gProGlyTyrHisAspIleValThrIleProAlaGlyAlaThrAsnIleG 753

    2323 ..........AGGAACCTTCACCCCCCCAAGGATGAAGGACCC..... 2355
                 ||||||   :::      :::::::::|||
     753 luValLysHisArgAsnGlnArgGlySerArgAsnAsnGlySerPheLeu 769

    2356 ......AAGGGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGG 2399
           ::::::   ::::::|||   ::::::   |||:::::::::
     770 AlaIleArgAlaAlaAspGlyThrTyrIleLeuAsnGlyAsnPheThrLe 786

    2400 TGCCATCCTGGTAGCAGCTATTGTCCTTGGGGGCACA..... 2436
          :::   |||         ::::::      ||||||
     786 uSerThrLeuGluGlnAspLeuThrTyrLysGlyThrValLeuArgTyrS 803

    2437 ..GGCTGGGGATTTAAAAATGTCAAGAAGAGAAGGTTCGATCCTACTCAG 2484
           |||   :::           :::   |||   |||:::|||   :::
     803 erGlySerSerAlaAlaLeuGluArgIleArgSerPheSerProLeuLys 819

    2485 CAA 2487
         :::
     820 Glu 820
```

[redacted]_name: SwissProt_39:ATS6_HUMAN

```
seq_documentation_block:
ID   ATS6_HUMAN     STANDARD;      PRT;   860 AA.
AC   Q9UKP5;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ADAM-TS 6 PRECURSOR (EC 3.4.24.-) (A DISINTEGRIN AND METALLOPROTEINASE
DE   WITH THROMBOSPONDIN MOTIFS 6) (ADAMTS-6) (ADAM-TS6).
GN   ADAMTS6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99395124; PubMed=10464288;
RA   Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
RT   "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
RT   Zinc Metalloproteases.";
RL   J. Biol. Chem. 274:25555-25563(1999).
CC   -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVELS IN PLACENTA AND BARELY
CC       DETECTABLE IN A NUMBER OF OTHER TISSUES.
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF140674; AAD56357.1; -.
DR   HSSP; P34179; 1IAG.
DR   MIM; 605008; -.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR000884; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00090; tsp_1; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS50092; TSP1; 1.
DR   PROSITE; PS00427; DISINTEGRINS; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Repeat; Extracellular matrix.
FT   SIGNAL        1     21       POTENTIAL.
FT   PROPEP       22    244       BY SIMILARITY.
FT   CHAIN       245    860       ADAM-TS 6.
FT   METAL       403    403       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    404    404       BY SIMILARITY.
FT   METAL       407    407       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       413    413       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN      453    509       DISINTEGRIN-LIKE.
FT   DOMAIN      510    566       TSP-TYPE 1 1.
FT   DOMAIN      567    668       CYS-RICH.
FT   DOMAIN      669    795       SPACER.
FT   DOMAIN      796    852       TSP-TYPE 1 2.
FT   DOMAIN       68     71       POLY-ARG.
FT   DOMAIN      662    665       POLY-GLY.
FT   CARBOHYD     99     99       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    172    172       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    222    222       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    234    234       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    676    676       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    843    843       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   860 AA;  97098 MW;  E57213015DECB2C5 CRC64;


alignment_scores:
           Quality:  302.00            Length:    817
             Ratio:  0.851               Gaps:     40
 Percent Similarity:  43.452   Percent Identity: 21.053

alignment_block:
 US-09-634-252A-2 x ATS6_HUMAN   ..

 Align seg 1/1  to: ATS6_HUMAN  from: 1  to: 860

    286 AATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAG..... 324
        :::|||:::   ::::::|||||||||:::      :::
     22 HisSerAspHisArgLeuSerTyrSerSerGlnGluGluPheLeuThrTy 38

    325 .....AAAGAAATCACACTGCCTTCAAGA..... 348
                  ::::::|||:::|||   |||
     38 rLeuGluHisTyrGlnLeuThrIleProIleArgValAspGlnAsnGlyA 55

    349 .....CTCATATATTACATCAACCAAGACTCGGAAAGCCCTTATCACGTT 393
             |||   :::   :::::::::|||::::::|||
     55 laPheLeuSerPheThrValLysAsnAspLysHisSer..... 67

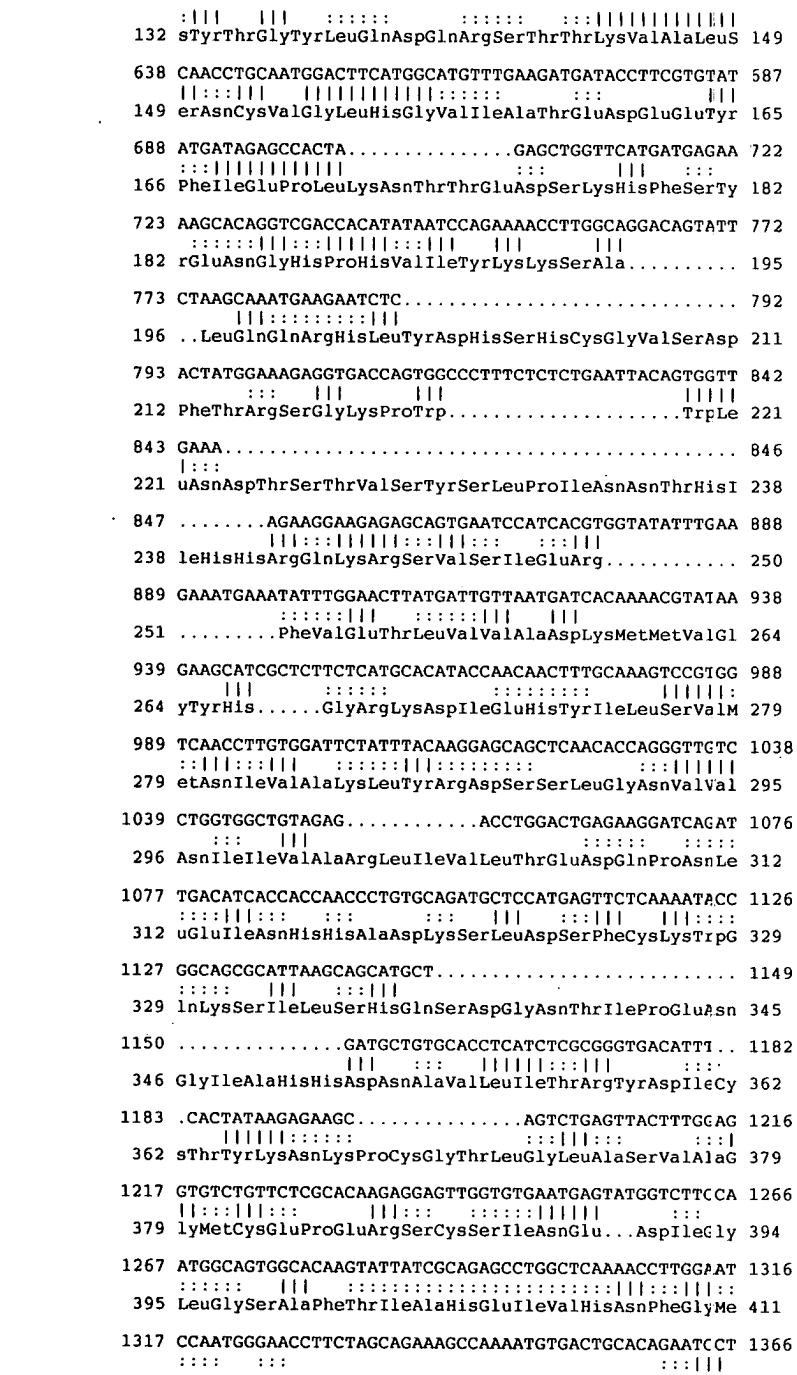
    394 CTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCT 443
              :::   |||   ::::::      :::          ::
     68 ......ArgArgArgArgSerMetAspProIleAspProGlnGlnAlaVa 82

    444 GGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTG 493
        :::::::      |||:::::::::|||:::|||:::   |||   |||:
     82 lSerLysLeuPhePheLysLeuSerAlaTyrGlyLysHisPheHisLeuA 99

    494 ACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCAC 543
        ::|||   ||||||:::   ::::::|||:::   ::::::      :::
     99 snLeuThrLeuAsnThrAspPheValSerLysHisPheThrValGluTyr 115

    544 TACGAAAATGGGAAACCACAGTACTCTAAGGGTGGA...GAGCACTGTTA 590
        :::   :::      ||||||::::::           ::::::|||::
    116 TrpGlyLysAspGlyProGlnTrpLysHisAspPheLeuAspAsnCysHi 132

    591 CTACCATGGAAGCATCAGAGGCGTCAAAGAC...TCCAAGGTGGCTCTGT 637
```


132 sTyrThrGlyTyrLeuGlnAspGlnArgSerThrThrLysValAlaLeuS 149
638 CAACCTGCAATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTAT 587
149 erAsnCysValGlyLeuHisGlyValIleAlaThrGluAspGluGluTyr 165
688 ATGATAGAGCCACTA.....GAGCTGGTTCATGATGAGAA 722
166 PheIleGluProLeuLysAsnThrThrGluAspSerLysHisPheSerTy 182
723 AAGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCAGGACAGTATT 772
182 rGluAsnGlyHisProHisValIleTyrLysLysSerAla..... 195
773 CTAAGCAAATGAAGAATCTC..... 792
196 ..LeuGlnGlnArgHisLeuTyrAspHisSerHisCysGlyValSerAsp 211
793 ACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTT 842
212 PheThrArgSerGlyLysProTrp.....TrpLe 221
843 GAAA..... 846
221 uAsnAspThrSerThrValSerTyrSerLeuProIleAsnAsnThrHisI 238
847 AGAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAA 888
238 leHisHisArgGlnLysArgSerValSerIleGluArg..... 250
889 GAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAA 938
251 PheValGluThrLeuValValAlaAspLysMetMetValGl 264
939 GAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGG 988
264 yTyrHis.....GlyArgLysAspIleGluHisTyrIleLeuSerValM 279
989 TCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTC 1038
279 etAsnIleValAlaLysLeuTyrArgAspSerSerLeuGlyAsnValVal 295
1039 CTGGTGGCTGTAGAG.....ACCTGGACTGAGAAGGATCAGAT 1076
296 AsnIleIleValAlaArgLeuIleValLeuThrGluAspGlnProAsnLe 312
1077 TGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACC 1126
312 uGluIleAsnHisHisAlaAspLysSerLeuAspSerPheCysLysTrpG 329
1127 GGCAGCGCATTAAGCAGCATGCT..... 1149
329 lnLysSerIleLeuSerHisGlnSerAspGlyAsnThrIleProGluAsn 345
1150 GATGCTGTGCACCTCATCTCGCGGGTGACATTT.. 1182
346 GlyIleAlaHisHisAspAsnAlaValLeuIleThrArgTyrAspIleCy 362
1183 .CACTATAAGAGAAGC.....AGTCTGAGTTACTTTGCAG 1216
362 sThrTyrLysAsnLysProCysGlyThrLeuGlyLeuAlaSerValAlaG 379
1217 GTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCA 1266
379 lyMetCysGluProGluArgSerCysSerIleAsnGlu...AspIleGly 394
1267 ATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAAT 1316
395 LeuGlySerAlaPheThrIleAlaHisGluIleValHisAsnPheGlyMe 411
1317 CCAATGGGAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCT 1366
411 tAsnHisAsp.....GlyIleGlyAsnSerC 420
1367 GGGGTGGCTGCATCATGGAGGAAACA.....GGGGTGTCCCATTCTCGA 1410
420 ysGlyArgLysValMetLysGlnGlnAsnTyrGlySerSerHis..... 434
1411 AAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGG 1460
435 TyrCysGluTyrGlnSerPhePheLeu..... 443
1461 AGGTGGAGCCTGCCTTTTCAACAGG...CCAACAAAGCTATTT..... 1500
444 ValCysLeuGlnSerArg***HisHisGlnLeuPheArgGluV 458
1500 1500
458 alCysArgGluLeuTrpCysLeuSerLysSerAsnArgCysValThrAsn 474
1501 GAGCCCACGGAATGTGGAAATGGATACGTGGAAGC 1535
475 SerIleProAlaAlaGluGlyThrLeuCysGlnThrGlyAsnIleGluLy 491
1536 TGGG.....GAGG 1543
491 sGlyTrpCysTyrGlnGlyAspCysValProPheGlyThrTrpProGlnS 508
1544 AGTGTGATTGTGGTTTTCATGTG...GAATGCTATGGATTATGCTGTAAG 1590
508 erIleAspGlyGlyTrpGlyProTrpSerLeuTrpGlyGluCysSerArg 524
1591 AAATGTTCC.....CTCTCCAACGGGGCTCACTGCAGCGACGGGCC 1631
525 ThrCysGlyGlyGlyValSerSerSerLeuArgHisCys...AspSerPr 540
1632 CTGC..... 1635
540 oAlaProSerGlyGlyGlyLysTyrCysLeuGlyGluArgLysArgTyrA 557
1636 TGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGTAT..... 1674
557 rgSerCysAsnThrAspProCysProLeuGlySerArgAspPheArgGlu 573
1675 ...GAATGCCGGGATGCTGTGAAC.....GAGTGTGA 1703
574 LysGlnCysAlaAspPheAspAsnMetProPheArgGlyLysTyrTyrAs 590
1704 TATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTC 1753
590 nTrpLysProTyrThrGlyGlyGlyValLysProCysAlaLeuAsnCysL 607
1754 ATAAGCAA.....GAC 1764
607 euAlaGluGlyTyrAsnPheTyrThrGluArgAlaProAlaValIleAsp 623
1765 GGATATGCATGCAATCAAAATCAG...GGCCGCTGCTACAATGGCGAGTG 1811
624 GlyThrGlnCysAsnAlaAspSerLeuAspIleCysIleAsnGlyGluCy 640
1812 CAAG.....ACCAGAGACAACC 1828
640 sLysHisValGlyCysAspAsnIleLeuGlySerAspAlaArgGluAspA 657
1829 AGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGC 1878
657 rgCysArgVal..... 660
1879 TATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGA 1928
661 CysGlyGlyGl 664
1929 TGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCT 1978
664 yGly.....SerThrCysAspAlaIleGluGlyPheP 675

```
    1979 TACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGT 2028
         ::                        ||||||   |||   ::::::
     675 he............AsnAspSerLeuProArgGlyGlyTyrMetGluVal 687

    2029 GAGATCATTCCAACTTCCTTCTACCAT...CAAGGCCGGGTGATTGACTG 2075
               ||||||   :::      |||   :::   |||   :::
     688 ValGlnIleProArgGlySerValHisIleGluValArgGluValAlaMe 704

    2076 CAGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTATGTAG 2125
          |||      :::::::::|||   ::::::   |||      |||:::
     704 tSerLysAsnTyrIleAlaLeuLysSerGluGlyAspAspTyrTyrIleA 721

    2126 AAGA....TGGAACGCCATGTGGCCC...GTCTATGATGTG......... 2159
         ::      |||:::      ||||||      :::||||||
     721 snGlyAlaTrpThrIleAspTrpProArgLysPheAspValAlaGlyThr 737

    2160 ...TTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCT 2206
            |||:::   :::   ||||||::::::      :::   |||   ||
     738 AlaPheHisTyrLysArgProThrAspGluProGluSerLeuGluAlaLe 754

    2207 G 2207
         |
     754 u 754
```

seq_name: SwissProt_39:ATS1_RAT

seq_documentation_block:

```
ID   ATS1_RAT       STANDARD;      PRT;   967 AA.
AC   Q9WUQ1;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ADAM-TS 1 PRECURSOR (EC 3.4.24.-) (A DISINTEGRIN AND METALLOPROTEINASE
DE   WITH THROMBOSPONDIN MOTIFS 1) (ADAMTS-1) (ADAM-TS1).
GN   ADAMTS1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RA   Liu X., Tu Y., Yin T., Johnstone E.M., Stephenson D.T., Clemens J.A.,
RA   Little S.P.;
RT   "Induction of a disintegrin and metalloprotease with the
RT   thrombospondin type I motif (ADAMTS).";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: CLEAVES AGGRECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC       INVOLVED IN ITS TURNOVER. HAS ANGIOGENIC INHIBITOR ACTIVITY (BY
CC       SIMILARITY). ACTIVE METALLOPROTEASE, WHICH MAY BE ASSOCIATED WITH
CC       VARIOUS INFLAMMATORY PROCESSES AS WELL AS DEVELOPMENT OF CANCER
CC       CACHEXIA. MAY PLAY A CRITICAL ROLE IN FOLLICULAR RUPTURE (BY
CC       SIMILARITY).
CC   -!- CATALYTIC ACTIVITY: CLEAVES AGGRECAN AT THE 1683-GLU-|-LEU-1684
CC       SITE, WITHIN THE CHONDROITIN SULFATE ATTACHEMENT DOMAIN.
CC   -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- INDUCTION: DOWN-REGULATED IN ENDOTHELIAL CELLS DERIVED FROM
CC       CIRRHOTIC LIVER.
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF149118; AAD34012.1; -.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR000884; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00090; tsp_1; 3.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS50092; TSP1; 3.
DR   PROSITE; PS00427; DISINTEGRINS; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Repeat; Extracellular matrix; Heparin-binding.
FT   SIGNAL        1    54       POTENTIAL.
FT   PROPEP       55   252       BY SIMILARITY.
FT   CHAIN       253   967       ADAM-TS 1.
FT   SITE        205   205       CYSTEINE SWITCH (POTENTIAL).
FT   METAL       401   401       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    402   402       BY SIMILARITY.
FT   METAL       405   405       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       411   411       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN      476   558       DISINTEGRIN-LIKE.
FT   DOMAIN      559   615       TSP-TYPE 1 1.
FT   DOMAIN      616   724       CYS-RICH.
FT   DOMAIN      725   857       SPACER.
FT   DOMAIN      858   907       TSP-TYPE 1 2.
FT   DOMAIN      908   967       TSP-TYPE 1 3.
FT   DOMAIN      194   198       POLY-ARG.
FT   CARBOHYD    547   547       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    720   720       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    764   764       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    782   782       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    945   945       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   967 AA;  105705 MW;  F93C864F6DCDB4CF CRC64;
```

alignment_scores:

```
            Quality:  298.50            Length:    841
              Ratio:  0.818               Gaps:     43
 Percent Similarity:  43.401   Percent Identity: 20.690
```

alignment_block:

US-09-634-252A-2 x ATS1_RAT ..

Align seg 1/1 to: ATS1_RAT from: 1 to: 967

```
    439 CATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCAT 488
        |||   :::   :::   :::::::::::||||||||||:::::::::
     74 HisAspSerThrThrLeuLeuArgLeuAspAlaPheGlyGlnGlnLeuHi 90

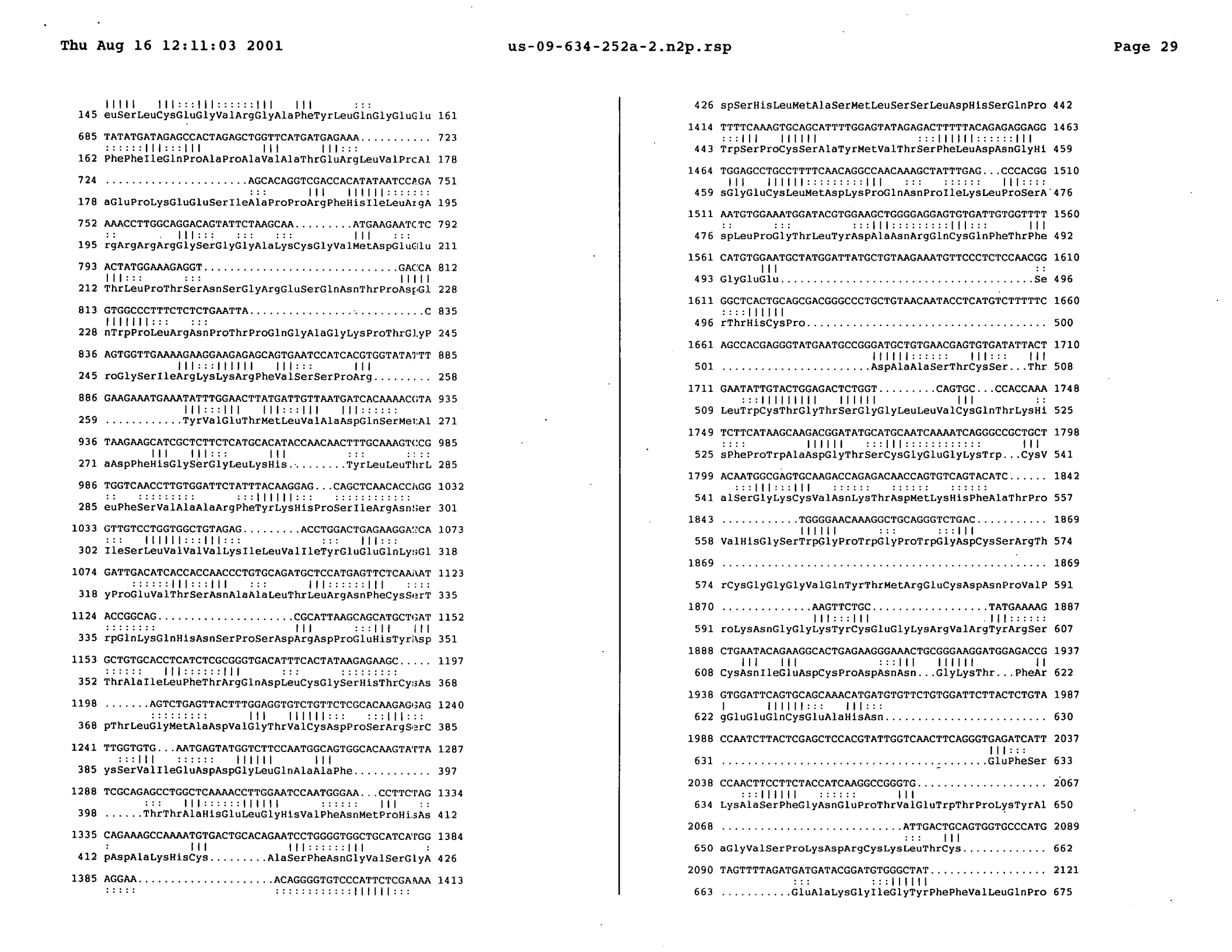
    489 TCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGA 538
         |||   |||      ::::::|||:::|||:::
     90 sLeuLysLeuGlnProAspSerGlyPheLeuAla..............  101

    539 TTCACTACGAAAATGGGAAACCACAGTAC.....................  567
                            |||   :::
    102 ....................ProGlyPheThrLeuGlnThrValGlyArg 111

    568 TCTAAGGGTGGAGAG.................................CA 584
        |||   |||:::|||                                  ||
    112 SerProGlySerGluAlaGlnHisLeuAspProThrGlyAspLeuAlaHi 128

    585 CTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTC 634
        ||||:::|||   |||:::::::::|||      :::|||   :::||||
    128 sCysPheTyrSerGlyThrValAsnGlyAspProSerSerAlaAlaAlaL 145

    635 TGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTG 684
```



```
    2122 ...GTAGAAGATGGAACGCCATGTGGCCCG......TCTATGATGTGTTT 2162
          |||   |||||||||||||||:::|||      :::   :::|||::
     676 LysValValAspGlyThrProCysSerProAspSerThrSerValCysVa 692

    2163 AGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCA. 2211
         ::::   :::|||::::::               .   :::::|||
     692 lGlnGlyGlnCysValLys.................AlaGlyCysAspA 703

    2212 .....CTCGATTCCAAGGGTAAA..............GTCTGTTCGGGC 2241
               :::||||||||||   |||              ||||||:::|||
     703 rgIleIleAspSerLysLysLysPheAspLysCysGlyValCysGlyGly 719

    2242 CATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGC 2291
         :::|||                 :::||||||       ||||||
     720 AsnGly..............SerThrCys.LysLysIleSer....... 728

    2292 AGGGACAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCA 2341
                                       ||||||   |||
     729 ............................GlyThrValThrSerThr 734

    2342 AGGATGAAGGACCCAAGGGTCCTAGTGCCACCAATCTCATAATAGGCTCC 2391
         |||
     735 Arg.......................................... 735

    2392 ATCGCTGGTGCCATCCTGGTAGCAGCTATTGTCCTTGGGGGCACAGGCTG 2441
                 |||::::::   ::::::::::::::   |||   |||
     736 .........ProGlyTyrHisAspIleValThrIleProAlaGlyAlaT 749

    2442 GGGATTTAAAAATG..................TCAAGAAGAGAAGGT 2470
         ::::::::::::                    ||||||:::.:::|||
     749 hrAsnIleGluValLysHisArgAsnProArgGlySerArgAsnAsnGly 765

    2471 TCGATCCTACTCAGCAAGGCC 2491
         |||:::|||      :::|||
     766 SerPheLeuAlaIleArgAla 772

seq_name: SwissProt_39:YE94_SCHPO

seq_documentation_block:
ID   YE94_SCHPO     STANDARD;      PRT;   512 AA.
AC   O13766;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   PROBABLE ZINC METALLOPEPTIDASE C17A5.04C PRECURSOR (EC 3.4.24.-).
GN   SPAC17A5.04C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RA   Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Wood V.;
RL   Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY ACT AS AN ATP-DEPENDENT ZINC METALLOPEPTIDASE
CC       (BY SIMILARITY).
CC   -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; Z98849; CAB11504.1; -.
DR   HSSP; P18619; 1FVL.
DR   MEROPS; M12.180; -.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF00200; disintegrin; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; FALSE_NEG.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Metal-binding; Signal.
FT   SIGNAL       1    15       POTENTIAL.
FT   CHAIN       16   512       PROBABLE ZINC METALLOPEPTIDASE C17A5.04C.
FT   METAL      229   229       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE   230   230       BY SIMILARITY.
FT   METAL      233   233       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL      239   239       ZINC (CATALYTIC) (PROBABLE).
SQ   SEQUENCE   512 AA;  56439 MW;  804BDA2333621DAB CRC64;


alignment_scores:
           Quality:  292.00            Length:    411
             Ratio:  1.264               Gaps:     21
 Percent Similarity:  56.204   Percent Identity: 25.061

alignment_block:
 US-09-634-252A-2 x YE94_SCHPO   ..

 Align seg 1/1  to: YE94_SCHPO  from: 1  to: 512

     868 CCATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTTATGATTGT 917
         |||      :::::::|||      ::::::   |||   :::   :::||
      56 ProValSerSerLeuPheProAlaLeuGlnThrLeuTrpIleGlyValVa 72

     918 TAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCA 967
         |   |||      :::|||      |||   :::|||:::   :::::::
      72 lAlaAsp...CysSerTyrValThrHisPheThrSerArgMetGluAlaL 88

     968 ACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAG 1017
         :::::         :::      :::   |||:::::::::|||::::::
      88 ysLysHisIlePheGlnGluPheGluGlyValSerThrLeuTyrGluAsp 104

    1018 CAGCTCAACACCAGGGTT..............GTCCTGGTGGCTGTAGA 1052
         ::::::|||   :::|||               :::|||   ::::::::
     105 SerPheAsnIleAsnValGlnIleHisSerLeuIleLeuProSerAlaHi 121

    1053 GACCTGGACTGAGAAG......GATCAGATTGACATCACCACCAACCCT. 1095
         :      :::   :::      |||:::   :::|||:::   :::|||
     121 sAspCysSerAlaAsnValValAspArgProGluIleSerMetSerProA 138

    1096 ...........GTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGC 1134
                     :::   |||:::   |||||||||:::
     138 rgIleSerIleGluGluLysLeuGluIlePheSerLysTrp......... 151

    1135 ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCA 1184
            |||   ::::::   :::   ::::::::::::   ::::::   ||
     152 ...LysTyrGluSerProGlyAsnAsnValPheGluAlaIleSerProHi 167

    1185 CTATAAGAGAAGC..................................... 1197
         |   :::::::|||
     167 sGluArgGluSerPheProSerGluProGlnValSerValLeuPheThrS 184

    1198 .....................AGTCTGAGTTACTTTGGAGGTGTCTGT 1224
                               ::::::|||:::|||:::   :::|||
     184 erSerValLysArgSerProHisGlyValSerTrpPheAlaThrIleCys 200

    1225 TCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTT...CCAATGGC 1271
         |||:::|||:::          ||||||:::   :::   |||:::::
     201 SerGluThrHis......IleGluAsnGluTrpHisValGlyProLeuSe 215

    1272 AGTGGCACAA.....................GTATTATCGCAGAGCCTGG 1300
```

```
     :|||::::::                        |||:::::::::::::::
  215 rValValSerAlaTyrProAsnAspArgLeuValValAlaHisGluIleG 232

 1301 CTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATGT 1350
      :::::   ||||||:::      :::          :::|||   :::|||
  232 lyHisIleLeuGlyLeuIleHisAsp......CysAsnLysLysSerCys 246

 1351 GAC.................TGCACAGAATCCTGGGGTGGCTGCATCAT 1382
                          |||      |||   :::   |||
  247 GlyAspHisSerGluAlaCysCysProLeuSerSerSerLeuCysAspAl 263

 1383 GGAGGAAACAGGGGTGTCCCAT.....................TCTC 1408
       :::|||      :::   :::                       :::
  263 aGlnGluLeuTyrIleMetAsnProSerAsnSerTyrThrTyrAlaAsnL 280

 1409 GAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGA 1458
        :::||||||   ||||||||||||:::   :::::::::::::::::
  280 euArgPheSerAspCysSerIleLeuGlnLeuHisSerLeuValGluLys 296

 1459 GGA......GGTGGAGCCTGCCTTTTCAACAGGCCAACAAAG......CT 1496
              :::   :::|||||||   :::::::|||::::::      ::
  297 LysTyrValSerLeuSerCysLeu...SerLysProSerGluLysSerVa 312

 1497 ATTTGAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGT 1546
      :::::::         ||||||||||||   ||||||   ||||||||||
  312 lLeuArgLeuGlyThrCysGlyAsnGlyIleValGluAspGlyGluGluC 329

 1547 GTGATTGTGGTTTTCATGTGGAATGCTATGGATTATGCTGT.....AAG 1590
      ||||||||||         |||      :::   ||||||      |||
  329 ysAspCysGlyGluAspCysGlu...AsnAsnProCysCysAspGlyLys 344

 1591 AAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGAC......GGGCCCTG 1634
        |||:::|||:::::::|||:::   |||:::|||            ||
  345 ThrCysLysLeuThrLysGlySerLeuCysAspAspGlnGlnAspAlaCy 361

 1635 CTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGG 1684
      ||||         :::|||   |||:::      |||      ||||||:
  361 sCys......TyrGlnCysHisPheLysAsnAlaGlyThrLeuCysArgG 376

 1685 ATGCTGTGAACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGT 1734
      ::::::::|||   ||||||      |||:::|||||||||   |||:::
  376 lnSerThrAsnProCysAspLysProGluPheCysThrGlyIleSerSer 392

 1735 CAGTGCCCACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAA 1784
      :::||||||   :::   :::   :::||||||      |||::::::::
  393 LysCysProValAspGluAsnTrpAspAspGlyArgIleCysGlnAspSe 409

 1785 T......CAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAACC 1828
      :      :::|||   |||   :::|||   |||   :::   ::::::|
  409 rLeuGlyMetGlySerCysAlaSerGlyValCysThrSerAlaSerArgG 426

 1829 AGTGTCAGTACATC......TGGGGAACAAAGGCTGCAGGGTCTGACAAG 1872
      ||||||:::   :::      :::::::::   ::::::   ||||||:::
  426 lnCysLysLysLeuThrAsnPheSerSerLeuSerCysHisSerAspSer 442

 1873 TTCTGCTATGAAAAGCTGAATACAGAAGGCACT 1905
               :::      |||   :::||||||
  443 CysLysValSerCysGlnAsnGluAspGlyThr 453

seq_name: SwissProt_39:ATS5_MOUSE

seq_documentation_block:
ID   ATS5_MOUSE     STANDARD;      PRT;   930 AA.
AC   Q9R001;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ADAM-TS 5 PRECURSOR (EC 3.4.24.-) (A DISINTEGRIN AND METALLOPROTEINASE
DE   WITH THROMBOSPONDIN MOTIFS 5) (ADAMTS-5) (ADAM-TS5) (AGGRECANASE-2)
DE   (ADMP-2) (IMPLANTIN).
GN   ADAMTS5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99395124; PubMed=10464288;
RA   Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
RT   "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
RT   zinc metalloproteases.";
RL   J. Biol. Chem. 274:25555-25563(1999).
CC   -!- FUNCTION: CLEAVES AGGRECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC       INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
CC       DESTRUCTION OF AGGRECAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN
CC       PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
CC   -!- CATALYTIC ACTIVITY: CLEAVES AGGRECAN AT THE 392-GLU-|-ALA-393
CC       SITE.
CC   -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED SPECIFICALLY IN THE PERI-
CC       IMPLANTATION PERIOD IN EMBRYO AND TROPHOBLAST AND AT LOW OR
CC       UNDECTABLE LEVEL THEREAFTER.
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF140673; AAD56356.1; -.
DR   MGD; MGI:1346321; Adamts5.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR000884; -.
DR   InterPro; IPR001590; -.
DR   Pfam; PF00090; tsp_1; 2.
DR   Pfam; PF01421; Reprolysin; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50092; TSP1; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS00427; DISINTEGRINS; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Repeat; Extracellular matrix.
FT   SIGNAL       1    21       POTENTIAL.
FT   PROPEP      22   261       POTENTIAL.
FT   CHAIN      262   930       ADAM-TS 5.
FT   SITE       209   209       CYSTEINE SWITCH (POTENTIAL).
FT   METAL      410   410       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE   411   411       BY SIMILARITY.
FT   METAL      414   414       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL      420   420       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN     485   566       DISINTEGRIN-LIKE.
FT   DOMAIN     567   623       TSP-TYPE 1 1.
FT   DOMAIN     624   731       CYS-RICH.
FT   DOMAIN     732   874       SPACER.
FT   DOMAIN     875   930       TSP-TYPE 1 2.
FT   DOMAIN      41    46       POLY-ALA.
FT   DOMAIN     257   261       POLY-ARG.
FT   CARBOHYD   498   498       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   728   728       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   802   802       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   807   807       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   930 AA;  101780 MW;  84DE84B26170D4DC CRC64;
```

```
alignment_scores:
            Quality:  290.50              Length:    842
              Ratio:  0.787                Gaps:     39
 Percent Similarity:  43.824   Percent Identity: 21.140

alignment_block:
 US-09-634-252A-2 x ATS5_MOUSE   ..

 Align seg 1/1  to: ATS5_MOUSE  from: 1  to: 930

     136 CTGCTTCTCGTCCTTCTCCTGCTGCCT...........CCGCTCGCCGC 173
         |||||||||:::|||||:|||||||                  |||||||
       8 LeuLeuLeuLeuLeuLeuLeuSerAlaSerCysLeuSerLeuAlaAl 24

     174 CTCGTCC.............................CGGCCCCGCG 190
         |:::|||                            :::|||:::|
      24 aAspSerProAlaAlaAlaProAlaGlnAspLysThrArgGlnProGlnA 41

     191 CCTGGGGGGCTGCTGCG...CCCAGCGCTCCGCATTGGAATGAAACTGCA 237
         ||   :::||||||||||   |||:::   |||:::   :::||||||
      41 laAlaAlaAlaAlaAlaGluProAspGlnProGlnGlyGluGluThrArg 57

     238 GAAAAA...AATTTGGGAGTCCTGGCAGATGAAGACAATACATTGCAACA 284
         |||:::   :::|||      ||||||                  |||::
      58 GluArgGlyHisLeuGlnProLeuAlaGly.............GlnAr 69

     285 GAATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAATCA 334
         ::::|||::::::
      69 gArgSerGlyGly................................... 73

     335 CACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGAAAGCCCT 384
                     |||::::::   |||:::|||      ::::::
      74 .............LeuValHisAsnIleAspGlnLeuTyrSerGlyGly 85

     385 TATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAGGC 434

      85 ................................................. 85

     435 TGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAAT 484
                 ::::::::::::::::   :::   |||   |||   :::|
      86 .........GlyLysValGlyTyrLeuValTyrAlaGlyGlyArgArgP 99

     485 TCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTG 534
         ||:::|||||||||      ::::::   :::   ::::::      :::
      99 heLeuLeuAspLeuGluArgAspAspThrValGlyAlaAlaGlySerIle 115

     535 GAGATTCACTACGAAAATGGGAAACCACAGTACTCTAAGGGTGGAGAGCA 584
            :::         :::|||            |||            ||
     116 ...ValThrAlaGlyGlyGlyLeuSerAlaSerSerGlyHisArgGlyHi 131

     585 CTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTC 634
         ||||:::|||:::|||::::::   |||        |||   ::::::
     131 sCysPheTyrArgGlyThrValAspGlySerProArgSerLeuAlaValP 148

     635 TGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTG 684
         :::::   |||:::||||||   |||:::|||
     148 heAspLeuCysGlyGlyLeuAspGlyPhePheAlaValLysHisAlaArg 164

     685 TATATGATAGAGCCACTA.......................GAGCTGGT 710
         |||   ::::::|||||                        |||   ::
     165 TyrThrLeuLysProLeuLeuArgGlySerTrpAlaGluTyrGluArgIl 181

     711 TCATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAA....... 753
         ::::   :::   |||:::          |||:::   ::::::
     181 eTyrGlyAspGlySerSerArgIleLeuHisValTyrAsnArgGluGlyP 198

     754 ....................................ACCTTGGCAGGA 765
                                             |||   |||:::
     198 heSerPheGluAlaLeuProProArgAlaSerCysGluThrProAlaSer 214

     766 ............CAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAG 803
                     |||   |||   ::::::   :::   :::   :::||
     215 ProSerGlyProGlnGluSerProSerValHisSerArgSerArgArgAr 231

     804 AGGTGACCAGTGGCCCTTTCTCTCTGAA..................... 831
         |:::         |||   |||   :::
     231 gSerAlaLeuAlaProGlnLeuLeuAspHisSerAlaPheSerProSerG 248

     832 ..............TTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAAT 867
                           |||.  :::||||||:::|||:::::::::
     248 lyAsnAlaGlyProGlnThrTrpTrpArgArgArgArgArgSerIleSer 264

     868 CCATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTTATGATTGT 917
            :::|||   :::                       ||||||:::::||
     265 ArgAlaArgGlnVal.....................GluLeuLeuLeuVa 274

     918 TAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCA 967
         |   |||   :::      :::   :::      :::      |||
     274 lAlaAspSerSerMetAlaArgMetTyrGlyArgGlyLeuGlnHisTyrL 291

     968 ACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAG 1017
                    |||:::::::|||   :::   |||      :::
     291 euLeuThrLeuAlaSerIleAlaAsnArgLeuTyrSerHisAlaSerIle 307

    1018 CAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAA 1067
         :::   :::   |||:::::::::|||   |||   :::   |||:::||
     308 GluAsnHisIleArgLeuAlaValValLysValValValLeuThrAspLy 324

    1068 GGAT...CAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGT 1114
         ||||   ::::::::::::::   |||   :::      |||   :::|
     324 sAspThrSerLeuGluValSerLysAsnAlaAlaThrThrLeuLysAsnP 341

    1115 TCTCAAAATACCGGCAGCGCATTAAGCAG..................... 1143
         ||   |||::::::::::::   :::|||
     341 heCysLysTrpGlnHisGlnHisAsnGlnLeuGlyAspAspHisGluGlu 357

    1144 CATGCTGATGCTGTGCACCTCATCTCGCGG......GTGACATTTCACTA 1187
         |||   ||||||:::   |||::::::|||      :::      |||::
     358 HisTyrAspAlaAlaIleLeuPheThrArgGluAspLeuCysGlyHisHi 374

    1188 TAAGAGAAGCAGTCTGAGTTAC......TTTGGAGGTGTCTGTTCTCGCA 1231
         ::::   :::::::|||:::            |||   :::||||||
     374 sSerCysAspThrLeuGlyMetAlaAspValGlyThrIleCysSerProG 391

    1232 CAAGAGGAGTTGGTGTG...AATGAGTATGGTCTTCCAATGGCAGTGGCA 1278
           |||:::   :::|||   ::::::   ||||||      |||
     391 luArgSerCysAlaValIleGluAspAspGlyLeuHisAlaAlaPhe... 406

    1279 CAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACC 1328
         ::::::::::::::::::::::::   ||||||::::::   :::
     407 ...ThrValAlaHisGluIleGlyHisLeuLeuGlyLeuSerHisAspAs 422

    1329 TTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGC. 1377
          |||:::                  |||   |||::::::|||
     422 pSerLys..............PheCysGluGluAsnPheGlyThrThrG 434

    1378 ...........ATCATGGAGGAA......ACAGGGGTGTCCCATTCTCGA 1410
                     :::|||::::::      |||:::::::::   |||:::
     434 luAspLysArgLeuMetSerSerIleLeuThrSerIleAspAlaSerLys 450

    1411 AAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGG 1460
            :::||||||||:::              :::|||||:::   ||
     451 ProTrpSerLysCysThrSerAlaThrIleThrGluPheLeuAspAspGl 467

    1461 AGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAG...CTATTTGAGCCCA 1507
         |   |||   ||||||::::::   |||   |||   ::::::   |||
     467 yHisGlyAsnCysLeuLeuAspLeuProArgLysGlnIleLeuGlyProG 484

    1508 CGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGT 1557
           |||      :::         :::|||   :::::::|||:::
     484 luGluLeuProGlyGlnThrTyrAspAlaThrGlnGlnCysAsnLeuThr 500

    1558 TTTCATGTGGAA.....................TGCTATGGATT 1580
         |||      |||                     |||      ||
     501 PheGlyProGluTyrSerValCysProGlyMetAspValCysAlaArgLe 517
```

```
 1581 ATGCTGT.....................................AAGAAAT 1594
        |   |||                                    ||||||
  517 uTrpCysAlaValValArgGlnGlyGlnMetValCysLeuThrLysLysL 534

 1595 GTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAAT 1644
                :::|||:::   |||:::   |||      |||   :::
  534 euProAlaValGluGlyThrProCysGlyLysGlyArgValCysLeuGln 550

 1645 ACCTCATGTCTT..................................... 1656
         :::|||:::
  551 GlyLysCysValAspLysThrLysLysLysTyrTyrSerThrSerSerHi 567

 1657 ..............TTTCAGCCACGAGGGTATGAATGCCGGGAT.... 1686
                     :::   |||   |||         |||:::
  567 sGlyAsnTrpGlySerTrpGlyProTrpGlyGlnCysSerArgSerCysG 584

  687 ..............GCTGTGAACGAGTGT..................... 1701
                     |||   ::::::|||
  584 lyGlyGlyValGlnPheAlaTyrArgHisCysAsnAsnProAlaProArg 600

 1702 GATATTACTGAATATTGTACTGGAGACTCTGGT................. 1734
        :::      :::||||||||||||      :::
  601 AsnSerGlyArgTyrCysThrGlyLysArgAlaIleTyrArgSerCysSe 617

 1735 .......CAGTGCCCACCAAATCTTCATAAGCAAGACGGATATGCATGCA 1777
              ||||||||||||
  617 rValThrProCysProProAsn........................... 624

 1778 ATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAAC 1827
              |||:::   :::                  |||   :::
  625 ..........GlyLysSerPhe.................ArgHisGlu 631

 1828 CAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTG 1877
        ||||||:::
  632 GlnCysGlu....................................... 634

 1878 CTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGG 1927
             |||   :::   ::::::      ||||||         |||
  635 ....AlaLysAsnGlyTyrGlnSerAspAlaLysGly......ValLysT 648

 1928 ATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTC 1977
              :::|||:::
  648 hrPheValGluTrpVal............................... 653

 1978 TTACTCTGTACCAATCTTACTCGAGCTCCACGT...ATTGGTCAACTTCA 2024
                                  |||:::     |||   |||
  654 .........................ProLysTyrAlaGlyValLeuPr 661

 2025 GGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTGACT 2074
        :::::::::          :::          :::|||
  661 oAlaAspValCysLysLeuThrCysArgAlaLysGly.............. 673

 2075 GCAGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTATGTA 2124
        :::|||   :::|||||||:::::                      |||
  674 ..ThrGlyTyrTyrValValPheSerProLys..............Val 684

 2125 GAAGATGGAACGCCATGTGGCCCG......TCTATGATGTGTTTAGATCG 2168
         |||||||||   |||   |||      :::   :::|||:::
  685 ThrAspGlyThrGluCysArgProTyrSerAsnSerValCysValArgGl 701

 2169 GAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGAT. 2217
        :::|||::::::      :::   ::::::   :::      |||:::
  701 yArgCysValArgThrGlyCysAspGlyIleIleGlySerLysLeuGlnT 718

 2218 ..TCCAAGGGTAAAGTCTGTTCGGGC 2241
        :::|||      ||||||:::|||
  718 yrAspLysCysGlyValCysGlyGly 726

seq_name: SwissProt_39:ATS8_HUMAN

seq_documentation_block:
ID   ATS8_HUMAN     STANDARD;      PRT;   890 AA.
AC   Q9UP79; Q9NZS0;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ADAM-TS 8 PRECURSOR (EC 3.4.24.-) (A DISINTEGRIN AND METALLOPROTEINASE
DE   WITH THROMBOSPONDIN MOTIFS 8) (ADAMTS-8) (ADAM-TS8) (METH-2) (METH-8).
GN   ADAMTS8 OR METH2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lung;
RX   MEDLINE=99367466; PubMed=10438512;
RA   Vazquez F., Hastings G., Ortega M.-A., Lane T.F., Oikemus S.,
RA   Lombardo M., Iruela-Arispe M.L.;
RT   "METH-1, a human ortholog of ADAMTS-1, and METH-2 are members of a new
RT   family of proteins with angio-inhibitory activity.";
RL   J. Biol. Chem. 274:23349-23357(1999).
RN   [2]
RP   SEQUENCE OF 195-440 FROM N.A.
RX   MEDLINE=20079168; PubMed=10610729;
RA   Georgiadis K.E., Hirohata S., Seldin M.F., Apte S.S.;
RT   "ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on
RT   mouse chromosome 9 and human chromosome 11.";
RL   Genomics 62:312-315(1999).
CC   -!- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES.
CC   -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: HIGLY EXPRESSED IN ADULT AND FETAL LUNG, LOWER
CC       EXPRESSION IN BRAIN, PLACENTA, HEART, STOMACH AND FETAL BRAIN AND
CC       KIDNEY.
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF060153; AAD48081.1; -.
DR   EMBL; AF175283; AAF25806.1; -.
DR   HSSP; P34179; 1IAG.
DR   MIM; 605175; -.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR000884; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00090; tsp_1; 2.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS50092; TSP1; 2.
DR   PROSITE; PS00427; DISINTEGRINS; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Repeat; Extracellular matrix; Heparin-binding.
FT   SIGNAL       1     27       POTENTIAL.
FT   PROPEP      28    214       BY SIMILARITY.
FT   CHAIN      215    890       ADAM-TS 8.
FT   METAL      364    364       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE   365    365       BY SIMILARITY.
FT   METAL      368    368       ZINC (CATALYTIC) (BY SIMILARITY).
```

```
FT   METAL       374    374       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN      439    526       DISINTEGRIN-LIKE.
FT   DOMAIN      527    583       TSP-TYPE 1 1.
FT   DOMAIN      584    690       CYS-RICH.
FT   DOMAIN      691    832       SPACER.
FT   DOMAIN      833    890       TSP-TYPE 1 2.
FT   DOMAIN      202    205       POLY-PRO.
FT   CARBOHYD    345    345       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    401    401       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    466    466       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    491    491       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    600    600       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    195    195       E -> R (IN REF. 2).
FT   CONFLICT    413    440       YLTELLDGGHGDCLLDAPGAALPLPTGL -> FSGCHLQGW
FT                                IHFKYLCKCVSELKCDLMP (IN REF. 2).
SQ   SEQUENCE   890 AA;  96671 MW;  57D70EE03D5739D3 CRC64;

alignment_scores:
            Quality:  280.00              Length:    992
              Ratio:  0.650                Gaps:     52
rcent Similarity:  43.448   Percent Identity: 21.371

nment_block:
-09-634-252A-2 x ATS8_HUMAN   ..

 Align seg 1/1  to: ATS8_HUMAN  from: 1  to: 890

    100 CCTGCCAGCGCCCCGGCCCGCACGCCGCCCTGCCGCCTGCTTCTCGTCCT 149
        ||||||   |||      |||      |||      |||||||||:::||
      3 ProAlaProAlaAlaProArgTrpLeuProPheLeuLeuLeuLeuLeLe 19

    150 TCTCCTGCTGCCT...CCGCTCGCCGCCTCGTCCCGGCCCCGCGCCTGGG 196
        ||||||||||||         :::|||   :::||||||   |||   |
     19 uLeuLeuLeuProLeuAlaArgGlyAlaProAlaArgProAlaAlaGlyG 36

    197 GGGCTGCTGCGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAAAAAAAT 246
        ||   |||:::
     36 lyGlnAlaSer..................................... 39

    247 TTGGGAGTCCTGGCAGATGAAGACAATACATTGCAACAGAATAGCAGCAG 296

     39 ................................................ 39

    297 TAATATCAGTTACAGCAATGCAATGCAGAAAGAAATCACACTGCCTTCAA 346
                                    |||:::::::::|||:::|
     40 ............................GluLeuValValProThrA 46

    347 GACTCATATATTACATCAACCAAGACTCGGAAAGCCCTTATCACGTTCTT 396
        |||||
     46 rgLeu........................................... 47

    397 GACACAAAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCTGGC 446
                                  :::::|||   :::|||||
     48 ........................ProGlySerAlaGlyGluLeuAl 55

    447 CCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACC 496
        |         ::::::::::::|||||||||:::   |||:::|||   |
     55 a.........LeuHisLeuSerAlaPheGlyLysGlyPheValLeuArgL 69

    497 TCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTAC 546
        ||      ::::::::::::|||:::   ::::::   :::|||:::
     69 euAlaProAspAspSerPheLeuAlaProGluPhe...LysIleGluArg 84

    547 GAAAATGGGAAACCACAGTACTCTAAGGGTGGAGAGCAC........... 585
           :::|||:::  :::   :::   ||||||||||:::
     85 LeuGlyGlySerGlyArg...AlaThrGlyGlyGluArgGlyLeuArgGl 100

    586 .TGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTC 634
         |||::::::   |||::::::::::|||      :::|||   :::|||:
    100 yCysPhePheSerGlyThrValAsnGlyGluProGluSerLeuAlaAlaV 117

    635 TGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTG 684

        ::|||   |||:::||||||   |||   |||      |||
    117 alSerLeuCysArgGlyLeuSerGlySerPheLeuLeuAspGlyGluGlu 133

    685 TATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACAGGTCG 734
        :::   |||:::|||                          |||   :::::
    134 PheThrIleGlnProGlnGlyAla.........GlyGlySerLeuAlaGl 147

    735 ACCACATATAATCCAGAAA............................... 753
        :||||||   :::|||:::
    147 nProHisArgLeuGlnArgTrpGlyProAlaGlyAlaArgProLeuProA 164

    754 .................ACCTTGGCAGGACAGTATTCTAAGCAAATGAAG 786
                           :::   :::|||:::   ::::::|||
    164 rgGlyProGluTrpGluValGluThrGlyGluGlyGlnArgGln...... 178

    787 AATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACA 836
                     ||||||||||||:::            :::|||   ::
    179 ............GluArgGlyAspHisGlnGluAspSerGluGluGluSe 191

    837 GTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCA................ 870
        :      ::::::   :::   |||   :::|||
    191 rGlnGluGluGluAlaGluGlyAlaSerGluProProProProLeuGlyA 208

    871 ...........TCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTT 909
                    ::::::   :::   :::|||   ::::::::::|||
    208 laThrSerArgThrLysArgPheValSerGluAlaArgPheValGluThr 224

    910 ATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGC 959
        ::::::|||   |||   :::                :::           ||
    225 LeuLeuValAlaAspAlaSerMetAlaAlaPheTyr.........GlyAl 238

    960 ACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTT 1009
        |      :::|||          ::::::::::::::::::      ||||
    238 aAspLeuGlnAsnHisIleLeuThrLeuMetSerValAlaAlaArgIleT 255

   1010 AC..............AAGGAGCAGCTCAACACCAGGGTTGTCCTGGTG 1044
        ||               |||::::::::::|||      ||||||   |||
    255 yrLysHisProSerIleLysAsnSerIleAsnLeuMetValValLysVal 271

   1045 GCTGTA.........GAGACCTGGACTGAGAAGGATCAGATTGACATCAC 1085
           :::         |||   |||                  ::::::::
    272 LeuIleValGluAspGluLysTrpGlyPro...........GluValSe 284

   1086 CACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGCA 1135
        :   |||            |||:::::|||   :::::::::::::|||:
    284 rAspAsnGlyGlyLeuThrLeuArgAsnPheCysAsnTrpGlnArgArgP 301

   1136 TTAAGCAG.....................CATGCTGATGCTGTGCACCTC 1164
        :::::|||                     |||   |||::::::   |||
    301 heAsnGlnProSerAspArgHisProGluHisTyrAspThrAlaIleLeu 317

   1165 ATCTCGCGGGTGACATTTCACTATAAGAGAAGC..............AG 1199
        ::::::|||   :::|||      :::::::::              ::
    318 LeuThrArgGlnAsnPheCysGlyGlnGluGlyLeuCysAspThrLeuGl 334

   1200 TCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGTGTG. 1248
        :::::::   :::|||   :::|||:::   :::::::::   :::|||
    334 yValAlaAspIleGlyThrIleCysAspProAsnLysSerCysSerValI 351

   1249 ..AATGAGTATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGC 1296
          ::::::   ||||||      |||                  ::::::
    351 leGluAspGluGlyLeuGlnAlaAla.................HisThr 361

   1297 CTGGCTCAAAACCTTGGA......ATCCAATGGGAACCTTCTAGCAGAAA 1340
        ||||||::::::||||||      ::::::         ::::::   ||
    362 LeuAlaHisGluLeuGlyHisValLeuSerMetProHisAspAspSerLy 378

   1341 GCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGC............A 1378
        ||||         |||||||:::   :::|||                  :
    378 sPro.........CysThrArgLeuPheGlyProMetGlyLysHisHisV 392

   1379 TCATGGAGGAAACAGGGGTG......TCCCATTCTCGAAAATTTTCAAAG 1422
        ::|||            |||      :::::::::      :::|||
```

```
 392 alMetAlaProLeuPheValHisLeuAsnGlnThrLeuProTrpSerPro 408

1423 TGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTG 1472
     ||||||   :::         ::::::|||:::   |||   |||   ||
 409 CysSerAlaMetTyrLeuThrGluLeuLeuAspGlyGlyHisGlyAspCy 425

1473 CCTTTTCAACAGGCCA................................ 1488
     ||||::::::   |||
 425 sLeuLeuAspAlaProGlyAlaAlaLeuProLeuProThrGlyLeuProG 442

1489 ................................ACAAAGCTATTTGAG 1503
                                     :::::::|||
 442 lyArgMetAlaLeuTyrGlnLeuAspGlnGlnCysArgGlnIlePheGly 458

 504 CCC......ACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTG 1547
     |||         :::|||   |||       :::::::|||   :::   ||
 459 ProAspPheArgHisCysProAsn......ThrSerAlaGlnAspValCy 473

1548 TGATTGTGGTTTTCATGTGGAATGCTAT..............GGATTAT 1582
     |           ::::::   |||:::                 | ||
 473 s.........AlaGlnLeuTrpCysHisThrAspGlyAlaGluProLeuC 487

1583 GC...TGTAAGAAATGTTCCCTC......TCCAACGGGGCTCACTGCAGC 1623
     ||      |||:::   ||||||      :::::::|||:::   |||:::
 487 ysHisThrLysAsnGlySerLeuProTrpAlaAspGlyThrProCysGly 503

1624 GACGGGCCCTGCTGTAACAATACCTCATGTCTT................. 1656
       |||      |||::::::   |||||||||
 504 ProGlyHisLeuCysSerGluGlySerCysLeuProGluGluGluValGl 520

1657 ................................TTTCAGCCACGAG 1669
                                     :::   |||   |
 520 uArgProLysProValValAspGlyGlyTrpAlaProTrpGlyProTrpG 537

1670 GGTATGAATGC.....................CGGGATGCTGTGAAC 1695
     ||   ||||||                    :::   :::   :::
 537 ly...GluCysSerArgThrCysGlyGlyGlyValGlnPheSerHisArg 552

1696 GAGTGTGATATTACTGAA.............TATTGTACTGGA.. 1725
     ||||||         |||             ||||||   |||
 553 GluCysLysAspProGluProGlnAsnGlyGlyArgTyrCysLeuGlyAr 569

1726 ..................GACTCTGGTCAGTGCCCACCAAATCTTC 1753
                         :::   :::|||||||||
 569 gArgAlaLysTyrGlnSerCysHisThrGluGluCysProPro..... 583

1754 ATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAAT 1803
           ||||||   :::   :::::::::|||   :::   ||||||
 584 ........AspGlyLysSerPheArgGluGlnGlnCysGluLysTyrAsn 597

1804 GGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAA 1853
     :::         |||      :::                |||      ||
 598 AlaTyrAsnTyrThrAspMetAspGlyAsnLeuLeuGlnTrpValProLy 614

1854 GGCTGCAGGG.........TCTGACAAG.........TTCTGC..... 1878
     |   ||||||          |||:::         ||||||
 614 sTyrAlaGlyValSerProArgAspArgCysLysLeuPheCysArgAlaA 631

1879 ......................TATGAAAAGCTGAATACAGAAGGCACT 1905
                           :::|||            :::||||||
 631 rgGlyArgSerGluPheLysValPheGluAlaLysValIleAspGlyThr 647

1906 ............................GAGAAGGGAAACTGCGGGAA 1925
                                 :::|||:::|||   ||
 648 LeuCysGlyProGluThrLeuAlaIleCysValArgGlyGlnCysValLy 664

1926 GGATGGA...GACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGTG 1972
     |   |||   |||:::   ::::::      :::            ||||
 664 sAlaGlyCysAspHisValValAspSerProArgLysLeuAspLysCysG 681

1973 GATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACTT 2022
     ||      :::|||   :::      ::::::   |||      ::::::
 681 ly......ValCysGlyGlyLysGlyAsnSerCysArg......LysVal 693

2023 CAGGGTGAGATCATTCCAACTTCCTTC...TACCATCAAGGCCGGGTGAT 2069
     :::|||:::::::   ||||||:::::::   |||::::::      :::||
 694 SerGlySerLeuThrProThrAsnTyrGlyTyrAsnAspIleValThrIl 710

2070 TGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCT 2119
     |   ::::::::::::::::::   :::   :::   :::      |||
 710 eProAlaGlyAlaThrAsnIleAspValLysGlnArgSerHisProGlyV 727

2120 ATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGG 2169
           :::||||||:::            ::::::         |||
 727 alGlnAsnAspGlyAsnTyrLeu......AlaLeuLysThrAlaAspGly 741

2170 AAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGAT.. 2217
     :::   |||            |||   :::|||:::          |||
 742 GlnTyrLeuLeuAsnGlyAsnLeuAlaIleSerAlaIleGluGlnAspIl 758

2218 ....TCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAAT.... 2259
           ||||||   :::   :::      |||      ::::::
 758 eLeuValLysGlyThrIleLeuLysTyrSerGlySerIleAlaThrLeuG 775

2260 ............................GAAGCCACCTGCATTTGTGAT 2280
                                 |||         :::
 775 luArgLeuGlnSerPheArgProLeuProGluProLeuThrValGlnLeu 791

2281 TTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGTTAGGAACCT 2330
     :::|||      |||                              :::::
 792 LeuThrValProGly.............................GluVa 798

2331 TCACCCCCCCAAG..................................... 2343
     :   |||||||||
 798 lPheProProLysValLysTyrThrPhePheValProAsnAspValAspP 815

2344 .....GATGAAGGACCCAAGGGTCCTAGTGCCACCAATCTCATAATAGGC 2388
           ::::::   |||      ::::::||||||:::|||
 815 heSerMetGlnSerSerLysGluArgAlaThrThrAsnIleIle...... 829

2389 TCCATCGCTGGTGCCATCCTGGTAGCAGCTATTGTCCTTGGG........ 2430
                    :::|||   |||      |||||||||
 830 .........GlnProLeuLeuHisAlaGlnTrpValLeuGlyAspTrpSe 843

2431 ..................GGCACAGGCTGGGGATTTAAAAATGTCAAGA 2461
                        |||:::||||||      ::::::|||:::
 843 rGluCysSerSerThrCysGlyAlaGlyTrpGlnArgArgThrValGluC 860

2462 AGAGAAGGTTCGATCCTACTCAGCAA 2487
       |||      ||||||:::   |||
 860 ysArg......AspProSerGlyGln 866
```

seq_name: SwissProt_39:ATS5_HUMAN

seq_documentation_block:

ID ATS5_HUMAN STANDARD; PRT; 930 AA.
AC Q9UNA0; Q9UKP2;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ADAM-TS 5 PRECURSOR (EC 3.4.24.-) (A DISINTEGRIN AND METALLOPROTEINASE
DE WITH THROMBOSPONDIN MOTIFS 5) (ADAMTS-5) (ADAM-TS5) (AGGRECANASE-2)
DE (ADMP-2) (ADAM-TS 11).
GN ADAMTS5 OR ADMP2 OR ADAMTS11.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=99367476; PubMed=10438522;
RA Abbaszade I., Liu R.-Q., Yang F., Rosenfeld S.A., Ross O.H.,
RA Link J.R., Ellis D.M., Tortorella M.D., Pratta M.A., Hollis J.M.,
RA Wynn R., Duke J.L., George H.J., Hillman M.C. Jr., Murphy K.,
RA Wiswall B.H., Copeland R.A., Decicco C.P., Bruckner R., Nagase H.,

```
RA   Ito Y., Newton R.C., Magolda R.L., Trzaskos J.M., Hollis G.F.,
RA   Arner E.C., Burn T.C.;
RT   "Cloning and characterization of ADAMTS11, an aggrecanase from the
RT   ADAMTS family.";
RL   J. Biol. Chem. 274:23443-23450(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20289799; PubMed=10830953;
RA   Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA   Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,
RA   Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,
RA   Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA   Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA   Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA   Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA   Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA   Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA   Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA   Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA   Lehrach H., Reinhardt R., Yaspo M.-L.;
RT   "The DNA sequence of human chromosome 21.";
     Nature 405:311-319(2000).
     [3]
     SEQUENCE OF 413-930 FROM N.A.
     TISSUE=Fetal brain;
RX   MEDLINE=99395124; PubMed=10464288;
RA   Hurskainen T.L., Hirohata S., Seldin M.F., Apte S.S.;
RT   "ADAM-TS5, ADAM-TS6, and ADAM-TS7, novel members of a new family of
RT   zinc metalloproteases.";
RL   J. Biol. Chem. 274:25555-25563(1999).
CC   -!- FUNCTION: CLEAVES AGGRECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC       INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
CC       DESTRUCTION OF AGGRECAN IN ARTHRITIC DISEASES. MAY PLAY A ROLE IN
CC       PROTEOLYTIC PROCESSING MOSTLY DURING THE PERI-IMPLANTATION PERIOD.
CC   -!- CATALYTIC ACTIVITY: CLEAVES AGGRECAN AT THE 392-GLU-|-ALA-393
CC       SITE.
CC   -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVEL, PRIMARILY IN PLACENTA
CC       BUT ALSO IN OTHER TISSUES, SUCH AS HEART AND BRAIN, AND ALSO
CC       CERVIX, UTERUS, BLADDER, ESOPHAGUS, RIB CARTILAGE,
CC       CHONDROBLASTOMA, FIBROUS TISSUE AND JOINT CAPSULE FROM AN
CC       ARTHRITIC PATIENT.
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF142099; AAD49577.1; -.
DR   EMBL; AP001698; BAA95504.1; -.
DR   EMBL; AP001697; BAA95503.1; -.
DR   EMBL; AF141293; AAF02493.1; -.
DR   MIM; 605007; -.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR000884; -.
DR   InterPro; IPR001590; -.
DR   Pfam; PF00090; tsp_1; 2.
DR   Pfam; PF01421; Reprolysin; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS50092; TSP1; 2.
DR   PROSITE; PS00427; DISINTEGRINS; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Repeat; Extracellular matrix.
FT   SIGNAL       1    16       POTENTIAL.
FT   PROPEP      17   261       POTENTIAL.
FT   CHAIN      262   930       ADAM-TS 5.
FT   SITE       209   209       CYSTEINE SWITCH (POTENTIAL).
FT   METAL      410   410       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE   411   411       BY SIMILARITY.
FT   METAL      414   414       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL      420   420       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN     485   566       DISINTEGRIN-LIKE.
FT   DOMAIN     567   623       TSP-TYPE 1 1.
FT   DOMAIN     624   731       CYS-RICH.
FT   DOMAIN     732   874       SPACER.
FT   DOMAIN     875   930       TSP-TYPE 1 2.
FT   DOMAIN      37    41       POLY-ALA.
FT   DOMAIN     257   261       POLY-ARG.
FT   CARBOHYD   498   498       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   728   728       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   802   802       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   807   807       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   138   138       A -> G (IN REF. 2).
FT   CONFLICT   614   614       R -> H (IN REF. 3).
FT   CONFLICT   692   692       P -> L (IN REF. 2).
SQ   SEQUENCE   930 AA;  101715 MW;  B64281502F28193B CRC64;


alignment_scores:
            Quality:  265.50              Length:    843
              Ratio:  0.727                 Gaps:     37
 Percent Similarity:  43.298   Percent Identity: 20.403

alignment_block:
 US-09-634-252A-2 x ATS5_HUMAN   ..

 Align seg 1/1  to: ATS5_HUMAN  from: 1  to: 930

    117 CCGCACGCCGCCCTGCCGCCTGCTTCTCGTCCTTCT..... 152
        |||   ||||||:::   |||||||:::   |||:::
     16 ProLeuAlaAlaValGlyProAlaAlaThrProAlaGlnAspLysAlaGl 32

    153 .......CCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCGCGCCTGG 195
              |||:::|||:::|||  ||||||:::   |||   |||
     32 yGlnProProThrAlaAlaAla.AlaAlaGlnProArgArgArgGlnGly 48

    196 ..............GGGGCTGCTGCGCCCAGCGCTCCGCATTGGAATGA 230
                          |||      |||:::   ||||||
     49 GluGluValGlnGluArgAlaGluProProGlyHisProHisProLeuAl 65

    231 AACTGCAGAAAAAAATTTGGGAGTCCTGGCAGATGAAGACAATACATTGC 280
             :::::::::   |||:::::::   :::   |||:::
     65 aGlnArgArgArgSerLysGlyLeuValGlnAsnIleAspGlnLeu.... 80

    281 AACAGAATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAA 330
                                  ||||||::::::
     81 ..........................TyrSerGlyGly............ 84

    331 ATCACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGAAAG 380

     84 ..................................................... 84

    381 CCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATA 430

     84 ..................................................... 84

    431 AGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCC 480
                    ::::::::::::::::   :::   |||   |||
     85 .............GlyLysValGlyTyrLeuValTyrAlaGlyGlyArg 96

    481 AAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTA 530
        :::|||:::|||||||||      ::::::::::::      :::   ::
```

```
 97 ArgPheLeuLeuAspLeuGluArgAspGlySerValGlyIleAlaGlyPh 113

531 TGTGGAGATTCACTACGAAAATGGGAAACCACAGTACTCTAAGGGTGGAG 580
    :|||             ::::::   |||     :::      :
113 eValProAlaGlyGlyGlyThrSerAlaProTrpArgHisArg......S 128

581 AGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTG 630
    ::|||||:::|||:::|||::::::   :::         |||   :::
128 erHisCysPheTyrArgGlyThrValAspAlaSerProArgSerLeuAla 144

631 GCTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATGATACCTT 680
    :::::::::   |||:::||||||   |||:::|||
145 ValPheAspLeuCysGlyGlyLeuAspGlyPhePheAlaValLysHisAl 161

681 CGTGTATATGATAGAGCCACTA.........GAGCTGGTTCATGATGAGA 721
        |||   ::::::||||||                ::::::::::||||
161 aArgTyrThrLeuLysProLeuLeuArgGlyProTrpAlaGluGluGluL 178

722 AA............AGCACAGGTCGACCACATATAATCCAG......... 750
    ||          :::   |||      :::|||::::::
178 ysGlyArgValTyrGlyAspGlySerAlaArgIleLeuHisValTyrThr 194

751 ..................AAAACCTTGGCAGGACAGTATTCTAAGCAAAT 782
                       ::::::|||      :::   |||   :::
195 ArgGluGlyPheSerPheGluAlaLeuProProArgAlaSerCysGluTh 211

783 GAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCC............. 819
            |||   |||      ::::::   |||
211 rProAlaSerThrProGluAlaHisGluHisAlaProAlaHisSerAsnP 228

820 .................TTTCTCTCTGAATTA.................. 834
                    :::   |||:::|||
228 roSerGlyArgAlaAlaLeuAlaSerGlnLeuLeuAspGlnSerAlaLeu 244

835 .....:.....................CAGTGGTTGAAAAGAAGGAAGAG 857
                                |||   :::||||||:::||
245 SerProAlaGlyGlySerGlyProGlnThrTrpTrpArgArgArgArgAr 261

858 AGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATTTGGAAC 907
    |::::::::::   :::|||   :::          .          ||||
261 gSerIleSerArgAlaArgGlnVal.....................GluL 271

908 TTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCAT 957
    ||:::::|||   |||   :::      :::   :::      :::
271 euLeuLeuValAlaAspAlaSerMetAlaArgLeuTyrGlyArgGlyLeu 287

958 GCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTAT 1007
       |||                  |||::::::|||   :::   |||
288 GlnHisTyrLeuLeuThrLeuAlaSerIleAlaAsnArgLeuTyrSerHi 304

1008 TTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCT 1057
       :::   :::   :::   |||:::::::::|||   |||   :::
304 sAlaSerIleGluAsnHisIleArgLeuAlaValValLysValValValL 321

1058 GGACTGAGAAGGATCAG...ATTGACATCACCACCAACCCTGTGCAGATG 1104
         :::|||||:::   ::::::::::::   |||   :::
321 euGlyAspLysAspLysSerLeuGluValSerLysAsnAlaAlaThrThr 337

1105 CTCCATGAGTTCTCAAAATACCGGCAGCGCATTAAGCAG........... 1143
     |||   :::|||   |||:::::::::::::   :::|||
338 LeuLysAsnPheCysLysTrpGlnHisGlnHisAsnGlnLeuGlyAspAs 354

1144 ..........CATGCTGATGCTGTGCACCTCATCTCGCGG......GTGA 1177
             |||   ||||||:::   |||:::::::|||       :::
354 pHisGluGluHisTyrAspAlaAlaIleLeuPheThrArgGluAspLeuC 371

1178 CATTTCACTATAAGAGAAGCAGTCTGAGTTAC......TTTGGAGGTGTC 1221
         |||::::::   ::::::|||:::            |||   :::
371 ysGlyHisHisSerCysAspThrLeuGlyMetAlaAspValGlyThrIle 387

1222 TGTTCTCGCACAAGAGGAGTTGGTGTG...AATGAGTATGGTCTTCCAAT 1268
     ||||||      |||:::   :::|||   ::::::   ||||||
388 CysSerProGluArgSerCysAlaValIleGluAspAspGlyLeuHisAl 404

1269 GGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCC 1318
       |||          :::::::::::::::::::::::::   ||||||::::
404 aAlaPhe......ThrValAlaHisGluIleGlyHisLeuLeuGlyLeuS 419

1319 AATGGGAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGG 1368
     ::   :::   |||:::                  |||   |||::::::
419 erHisAspAspSerLys..............PheCysGluGluThrPhe 430

1369 GGTGGCTGC............ATCATGGAGGAA......ACAGGGGTGTC 1400
     |||:::                :::|||::::::      |||:::::::
431 GlySerThrGluAspLysArgLeuMetSerSerIleLeuThrSerIleAs 447

1401 CCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTT 1450
     :   |||:::   :::||||||||||:::                 :::||||
447 pAlaSerLysProTrpSerLysCysThrSerAlaThrIleThrGluPheL 464

1451 TACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAG...CTA 1497
     ||:::   |||   |||   ||||||::::::   |||   |||   :::
464 euAspAspGlyHisGlyAsnCysLeuLeuAspLeuProArgLysGlnIle 480

1498 TTTGAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTG 1547
     :::   |||   |||      :::          :::|||   ::::::||
481 LeuGlyProGluGluLeuProGlyGlnThrTyrAspAlaThrGlnGlnCy 497

1548 TGATTGTGGTTTTCATGTGGAA.........................T 1570
     |:::      |||      |||                          |
497 sAsnLeuThrPheGlyProGluTyrSerValCysProGlyMetAspValC 514

1571 GCTATGGATTATGCTGT............................... 1587
     ||      |||   |||
514 ysAlaArgLeuTrpCysAlaValValArgGlnGlyGlnMetValCysLeu 530

1588 ...AAGAAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTG 1634
        ||||||           :::|||:::   |||:::   |||
531 ThrLysLysLeuProAlaValGluGlyThrProCysGlyLysGlyArgIl 547

1635 CTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGG 1684
      |||   :::   :::|||
547 eCysLeuGlnGlyLysCys............................. 553

1685 ATGCTGTGAACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGT 1734
          |||::::::                |||   |||::::::   |||
554 .....ValAspLysThrLysLysLysTyrTyrSerThrSerSerHisGly 568

1735 CAGTGCCCACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAA 1784
                                                      ||
569 ..............................................As 569

1785 TCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTC 1834
     |   |||
569 nTrpGlySer..................................... 572

1835 AGTACATCTGGGGAACA.................AAGGCTGCAGGGTCT 1866
            ||||||:::                  :::::::::|||:::
573 ........TrpGlySerTrpGlyGlnCysSerArgSerCysGlyGlyGly 586

1867 GACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAA 1916
        :::|||:::|||:::      |||:::   :::   ::::::::::
587 ValGlnPheAlaTyrArgHisCysAsnAsnProAlaProArgAsnAsnGl 603

1917 CTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGT 1966
     :            |||   |||   |||      |||
603 yArgTyrCysThrGlyLysArgAlaIleTyrArgSer............. 615

1967 TCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGT 2016
       |||::::::::::   |||                  |||      |||
616 ..CysSerLeuMetProCys.................ProProAsnGly 625

2017 CAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAA........ 2058
     :::                          ||||||   |||:::
626 Lys.......................SerPheArgHisGluGlnCysGl 634
```

```
2059 .....GGCCGGGTGATTGACTGCAGTGGTGCCCATGTA..... 2091
              |||         |||::::::|||:::   :::
 634 uAlaLysAsnGlyTyrGlnSerAspAlaLysGlyValLysThrPheValG 651

2092 ......................GTTTTAGATGATGATACG......... 2109
                           ||||||      |||:::
 651 luTrpValProLysTyrAlaGlyValLeuProAlaAspValCysLysLeu 667

2110 ..............GATGTGGGCTAT.....................GT 2123
                     :::|||||||                      ||
 668 ThrCysArgAlaLysGlyThrGlyTyrTyrValValPheSerProLysVa 684

2124 AGAAGATGGAACGCCATGTGGCCCG......TCTATGATGTGTTTAGATC 2167
     |   |||||||||   |||   |||      :::   :::|||:::
 684 lThrAspGlyThrGluCysArgProTyrSerAsnSerValCysValArgG 701

2168 GGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGAT 2217
       ||||||::::::      :::   ::::::   :::      |||:::
 701 lyLysCysValArgThrGlyCysAspGlyIleIleGlySerLysLeuGln 717

2218 ...TCCAAGGGTAAAGTCTGTTCGGGC 2241
        :::|||      ||||||:::|||
 718 TyrAspLysCysGlyValCysGlyGly 726

Query_name: SwissProt_39:ATS8_MOUSE

seq_documentation_block:
ID   ATS8_MOUSE     STANDARD;      PRT;   905 AA.
AC   P57110;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ADAM-TS 8 PRECURSOR (EC 3.4.24.-) (A DISINTEGRIN AND METALLOPROTEINASE
DE   WITH THROMBOSPONDIN MOTIFS 8) (ADAMTS-8) (ADAM-TS8) (METH-2).
GN   ADAMTS8.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20079168; PubMed=10610729;
RA   Georgiadis K.E., Hirohata S., Seldin M.F., Apte S.S.;
RT   "ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on
RT   mouse chromosome 9 and human chromosome 11.";
RL   Genomics 62:312-315(1999).
CC   -!- FUNCTION: HAS ANTI-ANGIOGENIC PROPERTIES (BY SIMILARITY).
CC   -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN ADULT LUNG AND HEART
CC       AND LOW EXPRESSION DURING MOUSE DEVELOPMENT.
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 TSP TYPE-1 DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF175282; AAF25805.1; -.
DR   MGD; MGI:1353468; Adamts8.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR000884; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR002870; -.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50092; TSP1; 2.
DR   PROSITE; PS00427; DISINTEGRINS; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Repeat; Extracellular matrix; Heparin-binding.
FT   SIGNAL        1     28       POTENTIAL.
FT   PROPEP       29    228       BY SIMILARITY.
FT   CHAIN       229    905       ADAM-TS 8.
FT   METAL       378    378       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    379    379       BY SIMILARITY.
FT   METAL       382    382       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       388    388       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN      453    541       DISINTEGRIN-LIKE.
FT   DOMAIN      542    598       TSP-TYPE 1 1.
FT   DOMAIN      599    705       CYS-RICH.
FT   DOMAIN      706    847       SPACER.
FT   DOMAIN      848    905       TSP-TYPE 1 2.
FT   CARBOHYD    415    415       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    480    480       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    506    506       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    615    615       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   905 AA;  98879 MW;  124D4132B33A0CAE CRC64;

alignment_scores:
            Quality:  265.00            Length:    925
              Ratio:  0.664               Gaps:     49
 Percent Similarity:  43.135   Percent Identity: 20.865

alignment_block:
 US-09-634-252A-2 x ATS8_MOUSE   ..

 Align seg 1/1  to: ATS8_MOUSE  from: 1  to: 905

    124 CCGCCCTGCCGCCTGCTTCTCGTCCTTCTCCTGCTGCCTCCGCTCGCCGC 173
        ||||||      |||||||||:::   |||      |||||||||:::
     11 ProProLeuLeuLeuLeuLeuGlnLeuProProProProLeuVal.. 26

    174 CTCGTCCCGGCCCCGCGCCTGGGGGGCTGCTGCGCCCAGCGCTCCGCATT 223
                                   ::::::||||||::::::|||
     27 .......................CysGlyAlaProAlaGlyProGly. 34

    224 GGAATGAAACTGCAGAAAAAAATTTGGGAGTCCTGGCAGATGAAGACAAT 273

     34 ................................................. 34

    274 ACATTGCAACAGAATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCA 323
                                                 :::::::|||:::
     35 ..................................ThrGlyAlaGlnAl 39

    324 GAAAGAAATCACACTGCCTTCAAGACTCATATATTACATCAACCAAGACT 373
         :::|||::::::::::|||:::||||||
     39 aSerGluLeuValValProThrArgLeu..... 48

    374 CGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAA 423

     48 ................................................. 48

    424 CATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTT 473
           :::::|||   :::||||||         |||:::::::::|||||
     49 ProGlySerAlaSerGluLeuAla.........PheHisLeuSerAlaPh 62

    474 CGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTT 523
        ||||:::   |||:::|||   |||      :::   :::::::|||:::
     62 eGlyGlnGlyPheValLeuArgLeuAlaProAspAlaSerPheLeuAlaP 79

    524 CTGATTAT...GTGGAGATTCACTACGAAAATGGGAAACCACAGTACTCT 570
          ::::::   :::|||         :::   |||      :::   :::
     79 roGluPheLysIleGlu.........ArgLeuGlyGlySerSerAlaAla 92
```

```
 571 AAGGGTGGAGAG............CACTGTTACTACCATGGAAGCATCAG 608
     |||||||||              |||:::::   |||::::::::
  93 AlaGlyGlyGluProGlyLeuArgGlyCysPhePheSerGlyThrValAs 109

 609 AGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATG 658
     :|||   ::::::|||   :::|||:::|||         |||      |
 109 nGlyGluArgGluSerLeuAlaAlaMetSerCysValAlaGlyTrpSerG 126

 659 GCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTG 708
     ||   |||                  :::   |||:::|||
 126 lySerPheLeuLeuAlaGlyGluGluPheThrIleGlnProGlnGlyAla 142

 709 GTTCATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAACCTT 758
                 |||      :::|||||||   :::|||:::
 143 .........GlyAspSerLeuAspGlnProHisArgLeuGlnArgTrpGl 156

 759 GGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAGGTG 808
     ||||||          :::   :::::::      :::      ::::
 156 yProGlyGln.........ArgArgGluAspProGlyLeuAlaAlaAlaG 170

 809 ACCAGTGGCCCTTTCTCTCTGAATTACAGTGG..................... 840
     ::   :::|||:::   :::   |||:::|||
 170 luValPheProLeuProGlnGlyLeuGluTrpGluValGluMetGlyAsn 186

 840 ..................................................... 840

 187 GlyGlnGlyGlnGluArgSerAspAsnGluGluAspArgLysGlnAspLy 203

 841 .........TTGAAAAGAAGGAAG......AGAGCAGTGAATCCA.... 870
             ||||||:::   :::       |||   |||   |||
 203 sGluGlyLeuLeuLysGluThrGluAspSerArgLysValProProProP 220

 871 ................TCACGTGGTATATTTGAAGAAATGAAATATTTG 903
                     |||:::   ::: * :::|||   :::::::::
 220 heGlySerLysThrArgSerLysArgPheValSerGluAlaArgPheVal 236

 904 GAACTTATGATTGTTAATGATCACAAAACG.........TATAAGAAGCA 944
     |||   ::::::|||   |||   :::            |||
 237 GluThrLeuLeuValAlaAspAlaSerMetAlaAlaPheTyrGlyThrAs 253

 945 TCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACC 994
         ::::::|||      |||                   |||:::::::
 253 pLeuGlnAsnHisIleLeuThr.................ValMetSerM 264

 995 TTGTGGATTCTATTTACAAGGAGCAG....................... 1020
     :::::      |||||||||:::
 264 etAlaAlaArgIleTyrLysHisProSerIleArgAsnSerValAsnIeu 280

1021 CTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGA 1070
     :::   ::::::|||::::::|||      |||   |||
 281 ValValValLysValLeuIleValGluLysGluArgTrpGlyPro..... 295

1071 TCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAA 1120
            :::::::::   |||            |||:::::|||   :
 296 .......GluValSerAspAsnGlyGlyLeuThrLeuArgAsnPheCysS 310

1121 AATACCGGCAGCGCATTAAG.....................CAGCATGCT 1149
     ::::::::::|||::::::                      :::|||
 310 erTrpGlnArgArgPheAsnLysProSerAspArgHisProGluHisTyr 326

1150 GATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGC.. 1197
     |||::::::   |||:::::::|||   :::|||      |||   :::
 327 AspThrAlaIleLeuPheThrArgGlnAsnPheCysGlyLysGlyGluGl 343

1198 .............AGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAA 1234
                  :::::::::      |||   :::|||:::      :
 343 nCysAspThrLeuGlyMetAlaAspValGlyThrIleCysAspProAspL 360

1235 GAGGAGTTGGTGTG...AATGAGTATGGTCTTCCAATGGCAGTGGCACAA 1281
     :::::   :::|||   ::::::   ||||||      |||
 360 ysSerCysSerValIleLysAspGluGlyLeuGlnAlaAla......... 373

1282 GTATTATCGCAGAGCCTGGCTCAAAACCTTGGA......ATCCAATGGGA 1325
```

```
                  :::||||||:::::::||||||      ::::::
  374 .........TyrThrLeuAlaHisGluLeuGlyHisValLeuSerMetPr 387

 1326 ACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCT 1375
      ::::::   ||||||         |||::::::   :::|||
  387 oHisAspAspSerLysPro.........CysValArgLeuPheGly.... 399

 1376 GCATCATGGAGGAAACAGGG.............................. 1395
                      |||
  400 ...........ProMetGlyLysTyrHisMetMetAlaProPhePheIle 412

 1396 ...GTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAG 1442
         |||:::   :::       :::|||   ||||||   :::
  413 HisValAsnLysThrLeuProTrpSerProCysSerAlaValTyrLeuTh 429

 1443 AGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAA 1492
       :::::::|||:::   |||   |||   ||||||::::::   ||||||:
  429 rGluLeuLeuAspAspGlyHisGlyAspCysLeuLeuAspAlaProThrS 446

 1493 AGCTATTTGAGCCCACGGAATGTGGA............AATGGATACGTG 1530
      ::::::::   |||          |||             :::
  446 erValLeu...ProLeuProThrGlyLeuProGlyHisSerThrLeuTyr 461

 1531 GAAGCTGGGGAGGAGTGTGATTGTGGTTTT..................... 1560
      |||      ::::::|||          |||
  462 GluLeuAspGlnGlnCysLysGlnIlePheGlyProAspPheArgHisCy 478

 1561 ..........CATGTGGAA......TGCTATGGATTATGCTGTAAGAAA. 1593
                  ||||||       |||      ||||||::::::
  478 sProAsnThrSerValGluAspIleCysValGlnLeuCysAlaArgHisA 495

 1594 ....................TGTTCCCTCTCCAACGGGGCT......... 1614
                          |||       :::||||||:::
  495 rgAspSerAspGluProIleCysHisThrLysAsnGlySerLeuLeuTrp 511

 1615 ............CACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATG 1652
                     |||:::   |||       |||   :::   |||||
  512 AlaAspGlyThrProCysGlyProGlyHisLeuCysLeuAspGlySerCy 528

 1653 TCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTG 1702
      |:::::::::                    :::|||:::   |||
  528 sValLeuLys..................GluAspValGluAsnProLysA 539

 1703 ATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTT 1752
        :::::::::          ||||||   |||         |||
  539 laValValAsp.........GlyAspTrpGlyProTrpArgPro...... 550

 1753 CATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGC.........CG 1793
                           |||::::::::::   |||         ::
  551 ............TrpGlyGlnCysSerArgThrCysGlyGlyGlyIleGl 563

 1794 CTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCT 1843
      :      |||   ||||||          ||||||
  563 nPheSerAsnArgGluCys.........AspAsnProMetProGlnAsnG 577

 1844 GGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAAT 1893
        |||       :::   |||   ::::::       |||::::::   |||
  577 lyGlyArgPheCysLeuGly...GluArgValLysTyrGlnSerCysAsn 592

 1894 ACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGAC......CG 1937
      ||||||                 :::|||       :::|||         ||
  593 ThrGlu..............GluCysProProAsnGlyLysSerPheAr 604

 1938 GTGGATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTA 1987
      |       ||||||:::|||:::::::::::::
  604 gGluGlnGlnCysGluLysTyrAsnAlaTyr.................... 614

 1988 CCAATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATT 2037
        |||   |||                  :::|||:::|||::::::::::
  615 ..AsnHisThr.................AspLeuAspGlyAsnPheLeu 624

 2038 CCAACTTCCTTCTACCATCAAGGC........................... 2061
                    ::::::|||
```

```
     625 GlnTrpValProLysTyrSerGlyValSerProArgAspArgCysLysLe 641

    2062 ..............................CGGGTGATTGACTGCAGTG 2080
                                      :::|||:::::::::::::
     641 uPheCysArgAlaArgGlyArgSerGluPheLysValPheGluAlaLys. 657

    2081 GTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTATGTAGAAGAT 2130
                                            |||   |||
     658 ........................................ValIleAsp 660

    2131 GGAACGCCATGTGGCCCGTCTATGATG......TGTTTAGATCGGAAGTG 2174
         ||||||   ||||||||||:::   :::      |||:::      :::||
     661 GlyThrLeuCysGlyProAspThrLeuSerIleCysValArgGlyGlnCy 677

    2175 CCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCA......CTCGATT 2218
         |::::::      :::   ::::::   :::   |||      ||||||:
     677 sValLysAlaGlyCysAspHisValValAlaSnSerProLysLysLeuAspL 694

    2219 CCAAGGGTAAAGTCTGTTCGGGCCATGGG...GTGTGTAGTAATGAAGCC 2265
         ::   |||   ||||||:::|||   |||   :::|||   :::   :::
     694 ysCysGly...ValCysGlyGlyLysGlyThrAlaCysArgLysIleSer 709

     266 ACCTGCATTTGTGATTTCACCTGGGCAGGGACAGAT.............. 2301
               :::      |||:::::::::   :::|||
     710 GlySerPheThrProPheSerTyrGlyTyrAsnAspIleValThrIlePr 726

    2302 .......TGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCC...CCCA 2341
                 ::::::      |||      ::::::   ||||||      :
     726 oAlaGlyAlaThrAsnIleAspValLysGlnArgSerHisProGlyValA 743

    2342 AGGATGAAGGACCCAAGGGTCCTAGTGCCACC.................. 2373
         ::::::::|||                  |||
     743 rgAsnAspGlySerTyrLeuAlaLeuLysThrAlaAsnGlyGlnTyrLeu 759

    2374 .........AATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGC 2414
                   ||||||   |||:::::::||         |||||||||
     760 LeuAsnGlyAsnLeuAlaIleSerAlaIleGluGlnAspIleLeuValLy 776

    2415 AGCTATTGTCCTTGGGGGCACAGGC 2439
          :::   :::|||      :::|||
     776 sGlyThrIleLeuLysTyrSerGly 784

seq_name: SwissProt_39:ATS3_HUMAN

seq_documentation_block:
ID   ATS3_HUMAN     STANDARD;      PRT;  1201 AA.
AC   O15072;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ADAM-TS 3 PRECURSOR (EC 3.4.24.-) (A DISINTEGRIN AND METALLOPROTEINASE
DE   WITH THROMBOSPONDIN MOTIFS 3) (ADAMTS-3) (ADAM-TS3) (FRAGMENT).
     ADAMTS3 OR KIAA0366.
     Homo sapiens (Human).
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97349984; PubMed=9205841;
RA   Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. VII.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 4:141-150(1997).
CC   -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC   -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS4.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB002364; BAA20821.1; -.
DR   MIM; 605011; -.
DR   MEROPS; M12.220; -.
DR   InterPro; IPR000884; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00090; tsp_1; 4.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50092; TSP1; 2.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
DR   PROSITE; PS00427; DISINTEGRINS; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Repeat; Extracellular matrix; Heparin-binding.
FT   NON_TER       1      1
FT   SIGNAL       <1      ?       POTENTIAL.
FT   PROPEP        ?    245       BY SIMILARITY.
FT   CHAIN       246   1201       ADAM-TS 3.
FT   METAL       394    394       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    395    395       BY SIMILARITY.
FT   METAL       398    398       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       404    404       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN      466    546       DISINTEGRIN-LIKE.
FT   DOMAIN      547    603       TSP-TYPE 1 1.
FT   DOMAIN      604    708       CYS-RICH.
FT   DOMAIN      709    840       SPACER.
FT   DOMAIN      841    898       TSP-TYPE 1 2.
FT   DOMAIN      899    961       TSP-TYPE 1 3.
FT   DOMAIN      962   1013       TSP-TYPE 1 4.
FT   DOMAIN      242    245       POLY-ARG.
FT   CARBOHYD     79     79       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    115    115       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    238    238       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    341    341       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    471    471       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    810    810       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    938    938       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1201 AA;  135113 MW;  D54EA92BD506A3AA CRC64;


alignment_scores:
            Quality:  256.00          Length:    871
              Ratio:  0.646               Gaps:     47
 Percent Similarity:  45.465   Percent Identity: 21.240

alignment_block:
 US-09-634-252A-2 x ATS3_HUMAN   ..

 Align seg 1/1  to: ATS3_HUMAN  from: 1  to: 1201

     277 TTGCAACAGAATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGAA 326
         ::::::   ::::::|||::::::   :::   :::|||      :::
      10 ValGluValArgThrSerAlaAspGlyGlnAlaGlyAsnGluGluMetVa 26

     327 AGAAATCACACTGCCTTCAAGA.................CTCATATATT 358
         :::|||   ||||||   :::                  |||:::
      26 lGlnIleAspLeuProIleLysArgTyrArgGluTyrGluLeuValThrP 43
```

```
 359 ACATCAACCAAGACTCGGAAAGCCCTTAT......CACGTTCTTGACACA 402
 43 roValSerThrAsnLeuGluGlyArgTyrLeuSerHisThrLeuSerAla 59

 403 AAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGC 452
 60 SerHisLysLysArgSerAlaArgAspValSerSerAsnProGluGlnLe 76

 453 AAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATAC 502
 76 uPhePheAsnIleThrAlaPheGlyLysAspPheHisLeuArgLeuLysP 93

 503 TGAACAATGGTTTGTTGTCT...TCTGATTATGTGGAGATTCAC..... 543
 93 roAsnThrGlnLeuValAlaProGlyAlaValValGluTrpHisGluThr 109

 544 .....TACGAAAATGG 554
 110 SerLeuValProGlyAsnIleThrAspProIleAsnAsnHisGlnProGl 126

 555 GAAACCACAGTAC.....TCTAAGGGTGGAGAGCACTGTT 589
 126 ySerAlaThrTyrArgIleArgLysThrGluProLeuGlnThrAsnCysA 143

 590 ACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCA 639
 143 laTyrValGlyAspIleValAspIleProGlyThrSerValAlaIleSer 159

 640 ACCTGCAATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTACAT 689
 160 AsnCysAspGlyLeuAlaGlyMetIleLysSerAspAsnGluGluTyrPh 176

 690 GATAGAGCCACTAGAGCTGGTTCATGAT...GAGAAAAGCACAGGTCGAC 736
 176 eIleGluProLeuGluArgGlyLysGlnMetGluGluGluLysGlyArgI 193

 737 CACATATAATCCAGAAAACCTTGGCAGGACAGTATTCT..... 774
 193 leHisValValTyrLysArgSerAlaValGluGlnAlaProIleAspMet 209

 775 .....AAGCAAATGAAGAATCT 791
 210 SerLysAspPheHisTyrArgGluSerAspLeuGluGlyLeuAspAspLe 226

 792 CACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGT 841
 226 uGlyThrValTyrGlyAsnIleHisGlnGlnLeuAsnGluThr.....M 241

 842 TGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAA 891
 241 etArgArgArgArgHisAla.....GlyGlu 249

 892 ATGAAATAT...TTGGAACTTATGATTGTTAATGATCACAAAACGTATAA 938
 250 AsnAspTyrAsnIleGluValLeuLeuGlyValAspAspSerValValAr 266

 939 GAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGG 988
 266 gPheHis.....GlyLysGluHisValGlnAsnTyrLeuLeuThrLeuM 281

 989 TCAACCTTGTGGATTCTATTTACAAGGAGCAG.....CTCAACACC 1029
 281 etAsnIleValAsnGluIleTyrHisAspGluSerLeuGlyValHisIle 297

1030 AGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGA 1079
 298 AsnValValLeuValArgMetIleMetLeuGlyTyrAlaLysSerIleSe 314

1080 CATCACCACC.....AACCCTGTGCAGATGCTCCATGAGTTCTCAAAAT 1123
 314 rLeuIleGluArgGlyAsnProSerArgSerLeuGluAsnValCysArgT 331

1124 ACCGGCAGCGCATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGG 1173
 331 rpAlaSerGln...GlnGlnArgSerAspLeuAsnHis..... 342

1174 GTGACATTTCACTATAAGAGAAGC.....AGTCTGAGTTACTTTGG 1214
 343 ...SerGluHisHisAspHisAlaIlePheLeuThrArgGlnAspPheGl 358

1215 A.....GGTGTCTGTTCTCGCA 1231
 358 yProAlaGlyMetGlnGlyTyrAlaProValThrGlyMetCysHisProV 375

1232 CAAGAGGAGTTGGTGTGAAT...GAGTATGGTCTTCCAATGGCAGTGGCA 1278
 375 alArgSerCysThrLeuAsnHisGluAspGlyPheSerSerAlaPhe... 390

1279 CAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACC 1328
 391 ...ValValAlaHisGluThrGlyHisValLeuGlyMetGluHisAspGl 406

1329 TTCTAGCAGAAAGCCAAAATGTGACTGC...ACAGAATCCTGGGGTGGCT 1375
 406 yGlnGlyAsnArg.....CysGlyAspGluThrAlaMetGlyS 419

1376 GCATCATGGAGGAA......ACAGGGGTGTCCCATTCTCGAAAATTTTCA 1419
 419 erValMetAlaProLeuValGlnAlaAlaPheHisArgTyrHisTrpSer 435

1420 AAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGC 1469
 436 ArgCysSerGlyGlnGluLeuLysArgTyrIleHis......SerTyrAs 450

1470 CTGCCTTTTCAACAGGCCA.....ACAAAGCTATTTGAGC 1504
 450 pCysLeuLeuAspAspProPheAspHisAspTrpProLysLeu......P 465

1505 CCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGT 1554
 465 roGluLeuProGlyIleAsnTyr...SerMetAspGluGlnCysArgPhe 480

1555 GGTTTTCATGTGGAATGCTATGGATTATGC..... 1584
 481 AspPheGlyValGly...TyrLysMetCysThrAlaPheArgThrPheAs 496

1585 ....TGTAAGAAA......TGTTCCCTCTCC..... 1605
 496 pProCysLysGlnLeuTrpCysSerHisProAspAsnProTyrPheCysL 513

1606 .....AACGGGGCTCACTGCAGCGACGGGCCC 1632
 513 ysThrLysLysGlyProProLeuAspGlyThrGluCysAlaAlaGlyLys 529

1633 TGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCG 1682
 530 TrpCysTyrLysGlyHisCysMetTrpLysAsnAlaAsnGlnGlnLysGl 546

1683 GGATGCTGTGAACGAGTGTGATATTACTGAATAT......TGTACTGGAG 1726
 546 nAspGly...AsnTrpGlySerTrpThrLysPheGlySerCysSerArgT 562

1727 ACTCTGGT.....CAGTGCCCACCAAATCTT 1752
 562 hrCysGlyThrGlyValArgPheArgThrArgGlnCysAsnAsnProMet 578

1753 CATAAGCAAGACGGATATGCATGC.....AATCAAAATCAGGGCCG 1793
 579 ProIleAsnGlyGlyGlnAspCysProGlyValAsnPheGluTyrGlnLe 595

1794 CTGCTACAATGGCGAGTGCAAGACC.....AGAGACAACC 1828
 595 uCysAsnThrGluGluCysGlnLysHisPheGluAspPheArgAlaGlnG 612

1829 AGTGTCAG..... 1836
```

```
     |||||||||
  612 lnCysGlnGlnArgAsnSerHisPheGluTyrGlnAsnThrLysHisHis 628

 1837 .....TACATCTGGGG 1847
                  :::::::::
  629 TrpLeuProTyrGluHisProAspProLysLysArgCysHisLeuTyrCy 645

 1848 AACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAG 1897
         :::          :::|||             ::::::|||      :
  645 sGlnSerLysGluThrGlyAspValAlaTyrMetLysGlnLeuValHisA 662

 1898 AAGGCACTGAG.....AAGGGA 1914
      ::||||||:::          :::|||
  662 spGlyThrHisCysSerTyrLysAspProTyrSerIleCysValArgGly 678

 1915 AACTGCGGGAAGGATGGA...GACCGGTGGATTCAGTGCAGCAAACATGA 1961
      :::|||   |||   |||   |||:::   |||      :::|||   ::
  679 GluCysValLysValGlyCysAspLysGluIleGlySerAsnLysValGl 695

 1962 TGTGTTCTGTGGATTCTTACTCTGT..... 1986
      :      ||||||      :::|||
  695 uAspLysCysGly......ValCysGlyGlyAspAsnSerHisCysArgT 710

 1987 ........ACCAATCTTACTCGAGCTCCACGT...ATTGGTCAACTTCAG 2025
                :::::::|||||||:::||||||   :::|||   |||:::
  710 hrValLysGlyThrPheThrArgThrProArgLysLeuGlyTyrLeuLys 726

 2026 GGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTGACTG 2075
                |||||||
  727 MetPheAspIleProPro..... 732

 2076 CAGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTATGTAG 2125
       :::::::   |||||||:::::::::::::|||
  733 .GlyAlaArgHisValLeuIleGlnGluAsp..... 742

 2126 AAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGC 2175
                 :::::::|||   ::::::::::::   :::::::::
  743 ...........GluAlaSerProHisIleLeuAlaIleLysAsnGlnAla 755

 2176 CTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCCAAGGG 2225
                    ||||||
  756 ThrGlyHisTyrIleLeuAsn..... 762

 2226 TAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATT. 2274
                  |||   |||      :::::::::   |||   |||
  763 .............GlyLysGlyGluGluAlaLysSerArgThrPheIleA 775

 2275 ..TGTGATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGTT 2322
                    |||          |||   :::|||:::|||   :::
  775 spLeuGlyValGluTrp.........AspTyrAsnIleGluAspAspIle 788

 2323 AGGAACCTTCACCCCCCCAAGGATGAAGGACCCAAGGGTCCTAGTGCCAC 2372
      ::::::|||||||            :::|||||||            :::
  789 GluSerLeuHis.........ThrAspGlyProLeuHisAspProValIl 802

 2373 CAATCTCATAATA 2385
      |||||||||
  802 eValLeuIleIle 806

seq_name: SwissProt_39:HR2_TRIFL

seq_documentation_block:
ID   HR2_TRIFL      STANDARD;      PRT;   201 AA.
AC   P20165;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   TRIMERELYSIN II (EC 3.4.24.53) (H2 METALLOPROTEINASE).
OS   Trimeresurus flavoviridis (Habu).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Viperidae; Crotalinae; Trimeresurus.
OX   NCBI_TaxID=88087;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Venom;
RX   MEDLINE=89380121; PubMed=2777746;
RA   Takeya H., Arakawa M., Miyata T., Iwanaga S., Omori-Satoh T.;
RT   "Primary structure of H2-proteinase, a non-hemorrhagic
RT   metalloproteinase, isolated from the venom of the habu snake,
RT   Trimeresurus flavoviridis.";
RL   J. Biochem. 106:151-157(1989).
RN   [2]
RP   X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX   MEDLINE=97063307; PubMed=8907175;
RA   Kumasaka T., Yamamoto M., Moriyama H., Tanaka N., Sato M.,
RA   Katsube Y., Yamakawa Y., Omori-Satoh T., Iwanaga S., Ueki T.;
RT   "Crystal structure of H2-proteinase from the venom of Trimeresurus
RT   flavoviridis.";
RL   J. Biochem. 119:49-57(1996).
CC   -!- FUNCTION: MAJOR VENOM NON-HEMORRHAGIC METALLOPROTEINASE.
CC   -!- CATALYTIC ACTIVITY: CLEAVAGE OF 3-ASN-|-GLN-4, 10-HIS-|-LEU-11
CC      AND 14-ALA-|-LEU-15 IN THE INSULIN B CHAIN, AND THE BOND Z-GLY-
CC      PRO-|-LEU-GLY-PRO IN A SMALL MOLECULE SUBSTRATE OF MICROBIAL
CC      COLLAGENASE.
CC   -!- COFACTOR: BINDS ONE ZINC ION.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC      METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
DR   PIR; JU0037; HYTV2.
DR   HSSP; P34179; 2AIG.
DR   MEROPS; M12.155; -.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR001590; -.
DR   Pfam; PF01421; Reprolysin; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hydrolase; Metalloprotease; Zinc; Venom.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   METAL       142    142       ZINC (CATALYTIC).
FT   ACT_SITE    143    143
FT   METAL       146    146       ZINC (CATALYTIC).
FT   METAL       152    152       ZINC (CATALYTIC).
FT   DISULFID    117    196
FT   DISULFID    158    180
FT   DISULFID    160    163
SQ   SEQUENCE   201 AA;  23014 MW;  31DFA648717ADC10 CRC64;


alignment_scores:
          Quality:  223.00            Length:    204
            Ratio:  1.729               Gaps:      6
 Percent Similarity:  63.235   Percent Identity: 28.431

alignment_block:
 US-09-634-252A-2 x HR2_TRIFL   ..

 Align seg 1/1  to: HR2_TRIFL  from: 1  to: 201

  895 AAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCA 944
      :::|||:::||||||   ||||||   ||||||      |||||||||::
    6 ArgTyrIleGluLeuAlaIleValValAspHisGlyMetTyrLysLysTy 22

  945 TCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACC 994
      ::::::::::           :::      :::   ::::::||||||
   22 rAsnGlnAsnSerAspLysIleLysValArgValHisGlnMetValAsnH 39

  995 TTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTG 1044
         ::::::::::::|||:::      ||||||      :::   |||
   39 isIleAsnGluMetTyrArg...ProLeuAsnIleAlaIleSerLeuAsn 54

 1045 GCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCC 1094
         ::::::   |||::::::||||||   |||   :::   :::
   55 ArgLeuGlnIleTrpSerLysLysAspLeuIleThrValLysSerAlaSe 71

 1095 TGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATT..... 1137
              |||:::::::|||::::::::::|||:::   :::
```

```
     71 rAsnValThrLeuGluSerPheGlyAsnTrpArgGluThrValLeuLeuL 88

   1138 ..AAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCAC 1185
          :::|||:::   |||::::::||||||::::::   ::::::::::::
     88 ysGlnGlnAsnAsnAspCysAlaHisLeuLeuThrAlaThrAsnLeuAsn 104

   1186 TATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAG 1235
          :::      :::|||:::|||      |||:::|||:::
    105 AspAsnThrIleGlyLeuAlaTyrLysLysGlyMetCysAsnProLysLe 121

   1236 AGGAGTTGGTGTGAATGAGTATGGTCTTCCA......ATGGCAGTGGCAC 1279
         :::||||||:::   :::         |||      :::   ||||||
    121 uSerValGlyLeuValGlnAspTyrSerProAsnValPheMetValAlaV 138

    280 AAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCT 1329
          ::::::::::::::|||::::::||||||||:::::::   :::
    138 alThrMetThrHisGluLeuGlyHisAsnLeuGlyMetGluHisAspAsp 154

   1330 TCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCAT 1379
                |||   ||||||   |||                :::|||||
    155 .........LysAspLysCysLysCysGlu............AlaCysIl 164

   1380 CATGGAGGAAACA...GGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCA 1426
         ||||::::::::::   :::   :::   |||:::   ||||||   ||||
    164 eMetSerAspValIleSerAspLysProSerLysLeuPheSerAspCysS 181

   1427 GCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTT 1476
         ||      :::|||:::   ||||||   :::   :::      |||:::
    181 erLysAsnAspTyrGlnThrPheLeuThrLysTyrAsnProGlnCysIle 197

   1477 TTCAACAGGCCA 1488
         :::|||   |||
    198 LeuAsnAlaPro 201

seq_name: SwissProt_39:FIBR_AGKCO

seq_documentation_block:
ID   FIBR_AGKCO     STANDARD;      PRT;   203 AA.
AC   P28891;
DT   01-DEC-1992 (Rel. 24, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   FIBROLASE (EC 3.4.24.72) (FIBRINOLYTIC PROTEINASE).
OS   Agkistrodon contortrix contortrix (Southern copperhead).
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
     Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
     Viperidae; Crotalinae; Agkistrodon.
     NCBI_TaxID=8713;
     [1]
RP   SEQUENCE.
RC   TISSUE=Venom;
RX   MEDLINE=93278288; PubMed=1304358;
RA   Randolph A., Chamberlain S.H., Chu H.L.C., Retzios A.D.,
RA   Markland F.S. Jr., Masiarz F.R.;
RT   "Amino acid sequence of fibrolase, a direct-acting fibrinolytic
RT   enzyme from Agkistrodon contortrix contortrix venom.";
RL   Protein Sci. 1:590-600(1992).
CC   -!- FUNCTION: THIS PROTEIN IS A ZINC PROTEASE FROM SNAKE VENOM THAT
CC       ACTS IN HEMORRHAGE. IT CLEAVES FIBRINOGEN.
CC   -!- CATALYTIC ACTIVITY: CLEAVAGE OF 14-ALA-|-LEU-15 IN INSULIN B CHAIN
CC       AND CLEAVAGE OF 413-LYS-|-LEU-414 IN ALPHA CHAIN OF FIBRINOGEN.
CC   -!- COFACTOR: BINDS ONE ZINC ION.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
DR   PIR; A41827; HYSNFA.
DR   HSSP; P34179; 2AIG.
DR   MEROPS; M12.133; -.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR001590; -.
DR   Pfam; PF01421; Reprolysin; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hydrolase; Metalloprotease; Zinc; Venom.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   METAL       143    143       ZINC (CATALYTIC) (PROBABLE).
FT   ACT_SITE    144    144       BY SIMILARITY.
FT   METAL       147    147       ZINC (CATALYTIC) (PROBABLE).
FT   METAL       153    153       ZINC (CATALYTIC) (PROBABLE).
FT   DISULFID    118    198
FT   DISULFID    158    165       BY SIMILARITY.
FT   DISULFID    160    182       BY SIMILARITY.
FT   VARIANT       1      1       MISSING (IN SOME % OF THE CHAINS).
FT   VARIANT     189    189       T -> E.
FT   VARIANT     192    192       T -> L.
SQ   SEQUENCE   203 AA;  22908 MW;  646EBE6F7F1EA191 CRC64;


alignment_scores:
            Quality:  222.50              Length:    204
              Ratio:  1.673                 Gaps:      6
 Percent Similarity:  65.196   Percent Identity: 26.961

alignment_block:
 US-09-634-252A-2 x FIBR_AGKCO   ..

 Align seg 1/1  to: FIBR_AGKCO  from: 1  to: 203

    895 AAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCA 944
        :::|||::::::|||:::||||||   ||||||:::          |||::
      7 ArgTyrValGlnLeuValIleValAlaAspHisArgMetAsnThrLysTy 23

    945 TCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACC 994
        ::::::::::            ::::::::::::   ::::::||||||
     23 rAsnGlyAspSerAspLysIleArgGlnTrpValHisGlnIleValAsnT 40

    995 TTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTG 1044
          :::::::::||||||:::      ||||||   :::   :::||||||
     40 hrIleAsnGluIleTyrArg...ProLeuAsnIleGlnPheThrLeuVal 55

   1045 GCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCC 1094
        :::::|||   |||:::::::::|||   |||   :::|||:::
     56 GlyLeuGluIleTrpSerAsnGlnAspLeuIleThrValThrSerValSe 72

   1095 TGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGC.........A 1135
           :::   |||   :::|||::::::::::|||:::
     72 rHisAspThrLeuAlaSerPheGlyAsnTrpArgGluThrAspLeuLeuA 89

   1136 TTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCAC 1185
          :::|||:::   |||   ::::::|||::::::   :::   |||
     89 rgArgGlnArgHisAspAsnAlaGlnLeuLeuThrAlaIleAspPheAsp 105

   1186 TATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAG 1235
                  :::|||:::|||   ||||||:::|||:::      ::
     106 GlyAspThrValGlyLeuAlaTyrValGlyGlyMetCysGlnLeuLysHi 122

   1236 AGGAGTTGGTGTG......AATGAGTATGGTCTTCCAATGGCAGTGGCAC 1279
        :::::::||||||      ::::::   ::::::   :::   ||||||
     122 sSerThrGlyValIleGlnAspHisSerAlaIleAsnLeuLeuValAlaL 139

   1280 AAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCT 1329
          ::::::::::::::|||::::::||||||||:::::::   :::
     139 euThrMetAlaHisGluLeuGlyHisAsnLeuGlyMetAsnHisAspGly 155

   1330 TCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCAT 1379
        :::            ::::|||   |||      :::      :::|||::
     156 Asn............GlnCysHisCysGlyAlaAsn......SerCysVa 166

   1380 CATGGAGGAA...ACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCA 1426
        :|||            :::   :::   |||:::   ||||||   ||||
     166 lMetAlaAlaMetLeuSerAspGlnProSerLysLeuPheSerAspCysS 183

   1427 GCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTT 1476
        ||      :::|||:::   ||||||      ::::::      |||:::
     183 erLysLysAspTyrGlnThrPheLeuThrValAsnAsnProGlnCysIle 199

   1477 TTCAACAGGCCA 1488
```

```
        :::|||:::|||
    200 LeuAsnLysPro 203

seq_name: SwissProt_39:ATS2_HUMAN

seq_documentation_block:
ID   ATS2_HUMAN     STANDARD;      PRT;  1211 AA.
AC   O95450;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ADAM-TS 2 PRECURSOR (EC 3.4.24.14) (A DISINTEGRIN AND
DE   METALLOPROTEINASE WITH THROMBOSPONDIN MOTIFS 2) (ADAMTS-2) (ADAM-TS2)
DE   (PROCOLLAGEN I/II AMINO-PROPEPTIDE PROCESSING ENZYME) (PROCOLLAGEN I
DE   N-PROTEINASE) (PC I-NP) (PROCOLLAGEN N-ENDOPEPTIDASE) (PNPI)
DE   (PROCOLLAGEN I/II AMINO-PROPEPTIDE PROCESSING ENZYME).
GN   ADAMTS2 OR PCINP OR PCPNI.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
     [1]
     SEQUENCE FROM N.A. (ISOFORMS LPNPI AND SPNPI).
     TISSUE=Skin;
     MEDLINE=99347935; PubMed=10417273;
RA   Colige A., Sieron A.L., Li S.-W., Schwarze U., Petty E.,
RA   Wertelecki W., Wilcox W., Krakow D., Cohn D.H., Reardon W.,
RA   Byers P.H., Lapiere C.M., Prockop D.J., Nusgens B.V.;
RT   "Human Ehlers-Danlos syndrome type VII C and bovine dermatosparaxis
RT   are caused by mutations in the procollagen I N-proteinase gene.";
RL   Am. J. Hum. Genet. 65:308-317(1999).
CC   -!- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR
CC       TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO
CC       PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN
CC       COLLAGEN BIOSYNTHESIS.
CC   -!- CATALYTIC ACTIVITY: CLEAVES THE N-PROPEPTIDE OF COLLAGEN CHAIN
CC       ALPHA-1(I) AT PRO-|-GLN AND OF ALPHA-1(II) AND ALPHA-2(I) CHAINS
CC       AT ALA-|-GLN.
CC   -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC   -!- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO
CC       COLLAGEN TYPE XIV (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; LPNPI (SHOWN HERE) AND SPNPI;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING. SPNPI HAS NO SIGNIFICANT N-
CC       PROCOLLAGEN PEPTIDASE ACTIVITY.
CC   -!- TISSUE SPECIFICITY: EXPRESSED AT HIGHT LEVEL IN SKIN, BONE, TENDON
CC       AND AORTA AND AT LOW LEVELS IN THYMUS AND BRAIN.
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- DISEASE: DEFECTS IN ADAMTS2 ARE THE CAUSE OF EHLERS-DANLOS
CC       SYNDROME TYPE VIIC (EDS-VIIC), A RECESSIVELY INHERITED DISORDER
CC       CHARACTERIZED CLINICALLY BY SEVERE SKIN FRAGILITY AND
CC       BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN INCOMPLETELY
CC       PROCESSED AT THE AMINO TERMINUS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC   -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AJ003125; CAA05880.1; -.
DR   MIM; 604539; -.
DR   MIM; 225410; -.
DR   MEROPS; M12.301; -.
DR   InterPro; IPR000884; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00090; tsp_1; 4.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50092; TSP1; 4.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
DR   PROSITE; PS00427; DISINTEGRINS; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Repeat; Collagen degradation; Extracellular matrix; Heparin-binding;
KW   Alternative splicing.
FT   SIGNAL       1     29       POTENTIAL.
FT   PROPEP      30    253       BY SIMILARITY.
FT   CHAIN      254   1211       ADAM-TS 2.
FT   METAL      408    408       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE   409    409       BY SIMILARITY.
FT   METAL      412    412       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL      418    418       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN     480    560       DISINTEGRIN-LIKE.
FT   DOMAIN     561    617       TSP-TYPE 1 1.
FT   DOMAIN     618    722       CYS-RICH.
FT   SITE       691    693       CELL ATTACHMENT SITE (POTENTIAL).
FT   DOMAIN     723    851       SPACER.
FT   DOMAIN     852    911       TSP-TYPE 1 2.
FT   DOMAIN     912    974       TSP-TYPE 1 3.
FT   DOMAIN     975   1030       TSP-TYPE 1 4.
FT   DOMAIN      40     43       POLY-ALA.
FT   DOMAIN     185    188       POLY-GLU.
FT   CARBOHYD   112    112       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   251    251       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   949    949       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   993    993       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1031   1031       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1098   1098       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1145   1145       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1150   1150       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   544    566       HCFKGHCIWLTPDILKRDGSWGA -> FRPGAVAHACYPS
FT                               TLGGQGRWIA (IN ISOFORM SPNPI).
FT   VARSPLIC   567   1211       MISSING (IN ISOFORM SPNPI).
SQ   SEQUENCE  1211 AA;  134722 MW;  BECEEF25C23CAD2D CRC64;


alignment_scores:
            Quality:  219.50            Length:    854
              Ratio:  0.581               Gaps:     48
 Percent Similarity:  44.262   Percent Identity: 22.131

alignment_block:
 US-09-634-252A-2 x ATS2_HUMAN   ..

 Align seg 1/1  to: ATS2_HUMAN  from: 1  to: 1211

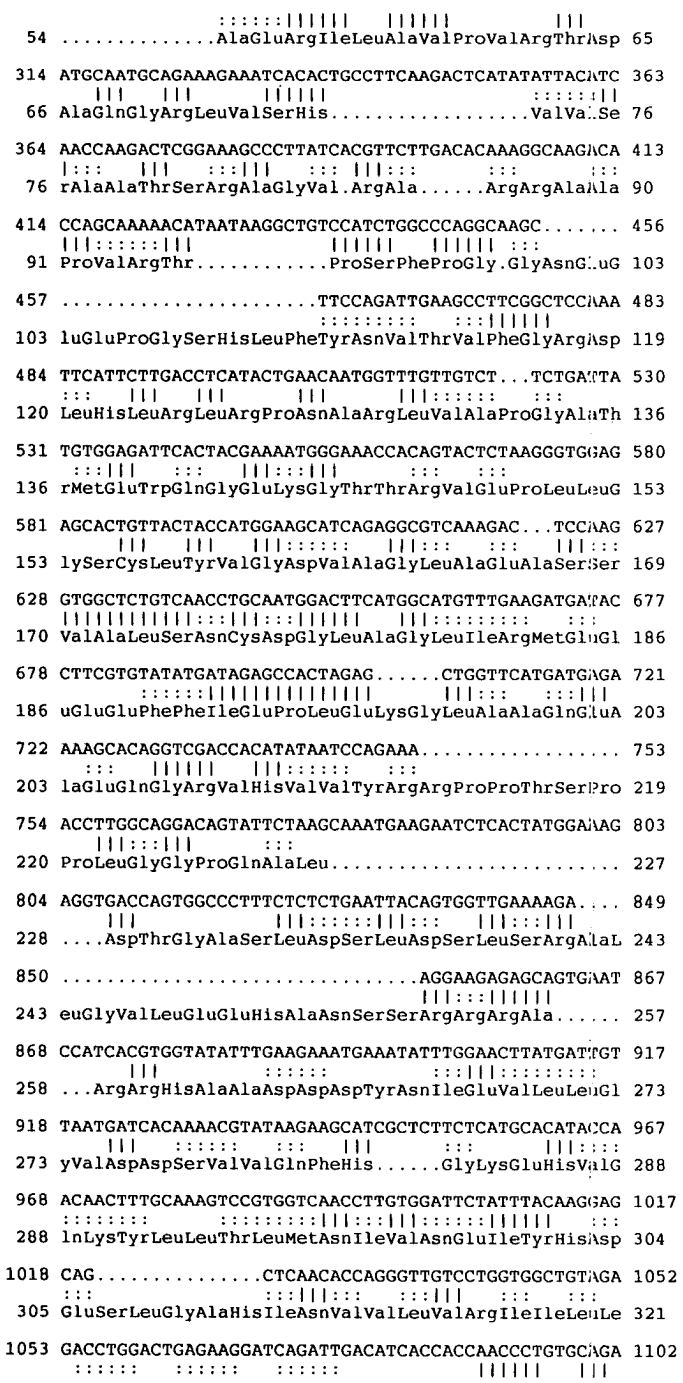
     100 CCTGCCAGCGCCCCGGCCCGCACGCCGCCCTGCCGCCTGCTTCTCGTCCT 149
         ||||||:::|||     |||              |||||||||:::||
       4 ProAlaGlyAlaAlaArgArgLeuLeuCysProAlaLeuLeuLeuLeLe 20

     150 TCTCCTGCTGCCTCCG......CTCGCCGCCTCGTCCCGGCCCCGCG... 190
         ||||||||||||||||      ||| ||||||   ||| . ||||||
      20 uLeuLeuLeuProProProLeuLeu.ProProProProProProAlaAsn 36

     191 .........................CCTGGGGGGCTGCTGCCGCCCAGC 213
                                  |||||||||   |||      ::
      37 AlaArgLeuAlaAlaAlaAlaAspProProGlyGlyProLeuGlyHisGl 53

     214 GCTCCGCATTGGAATGAAACTGCAGAAAAAAATTTGGGAGTCCTGGCAGA 263
         :
      53 y..................................................  53

     264 TGAAGACAATACATTGCAACAGAATAGCAGCAGTAATATCAGTTACAGCA 313
```


54 AlaGluArgIleLeuAlaValProValArgThrAsp 65
314 ATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACATC 363
66 AlaGlnGlyArgLeuValSerHis ValVa Se 76
364 AACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACA 413
76 rAlaAlaThrSerArgAlaGlyVal.ArgAla ArgArgAlaAla 90
414 CCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGC 456
91 ProValArgThr ProSerPheProGly.GlyAsnG uG 103
457 TTCCAGATTGAAGCCTTCGGCTCCAAA 483
103 luGluProGlySerHisLeuPheTyrAsnValThrValPheGlyArgAsp 119
484 TTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCT TCTGATTA 530
120 LeuHisLeuArgLeuArgProAsnAlaArgLeuValAlaProGlyAlaTh 136
531 TGTGGAGATTCACTACGAAAATGGGAAACCACAGTACTCTAAGGGTGGAG 580
136 rMetGluTrpGlnGlyGluLysGlyThrThrArgValGluProLeuLeuG 153
581 AGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGAC TCCAAG 627
153 lySerCysLeuTyrValGlyAspValAlaGlyLeuAlaGluAlaSerSer 169
628 GTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATGATAC 677
170 ValAlaLeuSerAsnCysAspGlyLeuAlaGlyLeuIleArgMetGluGl 186
678 CTTCGTGTATATGATAGAGCCACTAGAG CTGGTTCATGATGAGA 721
186 uGluGluPhePheIleGluProLeuGluLysGlyLeuAlaAlaGlnGluA 203
722 AAAGCACAGGTCGACCACATATAATCCAGAAA 753
203 laGluGlnGlyArgValHisValValTyrArgArgProProThrSerPro 219
754 ACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAG 803
220 ProLeuGlyGlyProGlnAlaLeu 227
804 AGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGA 849
228 AspThrGlyAlaSerLeuAspSerLeuAspSerLeuSerArgAlaL 243
850 AGGAAGAGAGCAGTGAAT 867
243 euGlyValLeuGluGluHisAlaAsnSerSerArgArgArgAla 257
868 CCATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTTATGATTGT 917
258 ArgArgHisAlaAlaAspAspAspTyrAsnIleGluValLeuLeuGl 273
918 TAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCA 967
273 yValAspAspSerValValGlnPheHis GlyLysGluHisValG 288
968 ACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAG 1017
288 lnLysTyrLeuLeuThrLeuMetAsnIleValAsnGluIleTyrHisAsp 304
1018 CAG CTCAACACCAGGGTTGTCCTGGTGGCTGTAGA 1052
305 GluSerLeuGlyAlaHisIleAsnValValLeuValArgIleIleLeuLe 321
1053 GACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAGA 1102
321 uSerTyrGlyLysSerMetSerLeuIleGluIleGlyAsnProSerGlnS 338
1103 TGCTCCATGAGTTCTCAAAATACCGGCAGCGCATTAAGCAGCATGCTGAT 1152
338 erLeuGluAsnValCysArgTrpAlaTyr LeuGlnGlnLysProAsp 353
1153 GCTGTGCAC CTCATCTCGCGGGTGAC 1178
354 ThrGlyHisAspGluTyrHisAspHisAlaIlePheLeuThrArgGlnAs 370
1179 ATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGA GGTG 1219
370 pPhe GlyProSerGlyMetGlnGlyTyrAlaProValThrGlyM 385
1220 TCTGTTCTCGCACAAGAGGAGTTGGTGTGAAT GAGTATGGTCTTCCA 1266
385 etCysHisProValArgSerCysThrLeuAsnHisGluAspGlyPheSer 401
1267 ATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAAT 1316
402 SerAlaPhe ValValAlaHisGluThrGlyHisValLeuGlyMe 416
1317 CCAATGGGAACCTTCTAGCAGAAAGCCAAAATGTGACTGC ACAGAAT 1363
416 tGluHisAspGlyGlnGlyAsnArg CysGlyAspGluV 429
1364 CCTGGGGTGGCTGCATCATGGAGGAA ACAGGGGTGTCCCATTCT 1407
429 alArgLeuGlySerIleMetAlaProLeuValGlnAlaAlaPheHisArg 445
1408 CGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAG 1457
446 PheHisTrpSerArgCysSerGlnGlnGluLeuSerArgTyrLeuHis 461
1458 AGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCA 1507
462 SerTyrAspCysLeuLeuAspAspProPheAlaHisAspTrpProA 477
1508 CG GAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGAT 1551
477 laLeuProGlnLeuProGlyLeuHisTyrSerMetAsnGluGlnCysArg 493
1552 TGTGGTTTTCATGTG 1566
494 PheAspPheGlyLeuGlyTyrMetMetCysThrAlaPheArgThrPheAs 510
1567 GAATGCTATGGATTATGCTGT 1587
510 pProCysLysGlnLeuTrpCysSerHisProAspAsnProTyrPheCysL 527
1588 AAGAAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCC 1632
527 ysThrLysLysGlyProProLeuAspGlyThrMetCysAlaProGlyLys 543
1633 TGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCG 1682
544 HisCysPheLysGlyHisCysIleTrpLeuThrProAspIleLeuLysAr 560
1683 GGATGCT GTGAACGAGTGTGATATTACTG 1711
560 gAspGlySerTrpGlyAlaTrpSerProPheGlySerCysSerArgThr 576
1712 AATATTGT ACTGGA GACTCTGGTCAGTGCCCACCAAAT 1749
577 CysGlyThrGlyValLysPheArgThrArgGlnCys AspAsn 590
1750 CTTCATAAGCAAGACGGA 1767
591 ProHisProAlaAsnGlyGlyArgThrCysSerGlyLeuAlaTyrAspPh 607
1768 TATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGA 1816
607 eGlnLeuCysSerArgGlnAsp CysProAspSerLeuAlaAspP 622

```
    1817 CCAGAGACAACCAGTGT..............................  1833
          |||::::::||||||
     622 heArgGluGluGlnCysArgGlnTrpAspLeuTyrPheGluHisGlyAsp  638

    1834 ...CAGTACATCTGG.................................  1845
            |||:::   |||
     639 AlaGlnHisHisTrpLeuProHisGluHisArgAspAlaLysGluArgCy  655

    1845 ..............................................  1845

     655 sHisLeuTyrCysGluSerArgGluThrGlyGluValValSerMetLysA  672

    1846 ..............GGAACAAAGGCTGCAGGGTCTGACAAGTTC......  1875
                      ||||||::::::::::   :::|||   |||
     672 rgMetValHisAspGlyThrArgCysSerTyrLysAspAlaPheSerLeu  688

    1876 TGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCGGGAA  1925
         |||                                :::|||:::|||   ||
     689 CysVal............................ArgGlyAspCysArgLy  696

    1926 GGATGGA...GACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGTG  1972
         |   |||   |||      |||      ||||||::::::      ||||
     696 sValGlyCysAspGlyValIleGlySerSerLysGlnGluAspLysCysG  713

    1973 GATTC..........................TTACTCTGTACCAATCTT  1995
         ||                            ::::::      ::::::
     713 lyValCysGlyGlyAspAsnSerHisCysLysValValLysGlyThrPhe  729

    1996 ACTCGAGCTCCA...CGTATTGGTCAACTTCAGGGTGAGATCATTCCAAC  2042
         ||||||:::|||   :::   |||   ::::::       ||||||::
     730 ThrArgSerProLysLysHisGlyTyrIleLysMetPheGluIleProAl  746

    2043 TTCCTTCTACCAT...CAAGGCCGGGTGATTGACTGCAGTGGTGCCCATG  2089
         ::::      |||        :::   :::|||:::::::::   |||:
     746 aGlyAlaArgHisLeuLeuIleGlnGluValAspAlaThrSerHisHisL  763

    2090 TA............................GTTTTAGATGATGATACG  2109
         ::                            :::|||:::::::::::
     763 euAlaValLysAsnLeuGluThrGlyLysPheIleLeuAsnGluGluAsn  779

    2110 GATGTG 2115
         ||||||
     780 AspVal 781
```

```
seq_name: SwissProt_39:HRT2_CRORU

seq_documentation_block:
ID   HRT2_CRORU     STANDARD;      PRT;   202 AA.
AC   P20897;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
     01-NOV-1995 (Rel. 32, Last annotation update)
     HEMORRHAGIC METALLOPROTEINASE HT-2 (EC 3.4.24.42) (RUBERLYSIN)
     (HEMORRHAGIC TOXIN II).
     Crotalus ruber ruber (Red diamond rattlesnake).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Viperidae; Crotalinae; Crotalus.
OX   NCBI_TaxID=8736;
RN   [1]
RP   SEQUENCE, AND DISULFIDE BONDS.
RC   TISSUE=Venom;
RX   MEDLINE=91185324; PubMed=2081731;
RA   Takeya H., Onikura T., Sugihara H., Iwanaga S.;
RT   "Primary structure of a hemorrhagic metalloproteinase, HT-2, isolated
RT   from the venom of Crotalus ruber ruber.";
RL   J. Biochem. 108:711-719(1990).
CC   -!- FUNCTION: THIS PROTEIN IS A ZINC PROTEASE FROM SNAKE VENOM THAT
CC       ACTS IN HEMORRHAGE BY CLEAVING FIBRINOGEN.
CC   -!- CATALYTIC ACTIVITY: CLEAVAGE OF 10-HIS-|-LEU-11, 14-ALA-|-LEU1-5,
CC       16-TYR-|-LEU-17 AND 23-GLY-|-PHE-24 BONDS OF THE B CHAIN OF
CC       INSULIN; HIS-|-PRO, PRO-|-PHE, AND TRP-|-SER OF ANGIOTENSIN I; AND
CC       GLY-|-PHE OF MET ENKEPHALIN.
CC   -!- COFACTOR: BINDS ONE ZINC ION.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
DR   PIR; JX0139; HYRSR.
DR   HSSP; P34179; 2AIG.
DR   MEROPS; M12.150; -.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR001590; -.
DR   Pfam; PF01421; Reprolysin; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hydrolase; Metalloprotease; Zinc; Venom.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   METAL       142    142       ZINC (CATALYTIC) (PROBABLE).
FT   ACT_SITE    143    143       BY SIMILARITY.
FT   METAL       146    146       ZINC (CATALYTIC) (PROBABLE).
FT   METAL       152    152       ZINC (CATALYTIC) (PROBABLE).
FT   DISULFID    117    197
FT   DISULFID    157    164
SQ   SEQUENCE   202 AA;  23320 MW;  5AF5D29318A61A9B CRC64;


alignment_scores:
            Quality:  212.50              Length:    205
              Ratio:  1.551                 Gaps:      7
 Percent Similarity:  66.829   Percent Identity: 26.341

alignment_block:
 US-09-634-252A-2 x HRT2_CRORU   ..

 Align seg 1/1  to: HRT2_CRORU  from: 1  to: 202

     898 TATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCG  947
         |||:::||||||:::::|||   |||||||:::   :::   |||:::::
       7 TyrIleGluLeuValValValAlaAspHisArgMetPheMetLysTyrAs   23

     948 CTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTG  997
         :|||:::           ::::::   :::   ::::::|||||::::
      23 nSerAspLeuAsnThrIleArgThrArgValHisGluIleValAsnPheI   40

     998 TGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCT  1047
         ::::::::::|||:::   :::||||||   ||||||   |||:::
      40 leAsnGluPheTyrArg...SerLeuAsnIleArgValSerLeuThrAsp   55

    1048 GTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGT  1097
         :::|||   |||::::::::::|||   |||   :::   ::::::
      56 LeuGluIleTrpSerAspGlnAspPheIleThrValGlnSerSerAlaLy   72

    1098 GCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATT.........  1137
          :::   ||||||:::|||::::::::::|||:::   :::
      72 sAsnThrLeuHisSerPheGlyGluTrpArgLysSerValLeuLeuAsnA   89

    1138 ..AAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCAC  1185
           |||:::|||   |||   ::::::|||::::::   :::::::::
      89 rgLysArgHis...AspAsnAlaGlnLeuLeuThrAlaIleValLeuAsp   104

    1186 TATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAG  1235
                    :::|||:::|||:::::::::::|||          ::
     105 AspTyrThrLeuGlyLeuAlaTyrLeuAsnSerMetCysHisProArgAs   121

    1236 AGGAGTTGGTGTGAATGAGTATGGTCTTCCA......ATGGCAGTGGCAC  1279
         ::::||||||:::   :::         |||       :::   ::::::
     121 nSerValGlyLeuIleGlnAspHisSerProIleAsnLeuLeuMetGlyV   138

    1280 AAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCT  1329
           ::::::::::::::|||::::::||||||||:::::   :::
     138 alThrMetAlaHisGluLeuGlyHisAsnLeuGlyMetGluHisAspGly   154

    1330 TCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCAT  1379
         :::                  ||||||   :::::   :::   |||||
     155 Lys.................AspCysLeuArgGlyAlaSerLeuCysIl   165
```

```
    1380 CATGGAGGAAACAGGGGTGTCCCATTCTCGA......AAATTTTCAAAGT 1423
         ||||   :::   |||::::::   :::|||      :::||||||   :
     165 eMet...ArgProGlyLeuThrProGlyArgSerTyrGluPheSerAspA 181

    1424 GCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGC 1473
         ::|||:::      |||:::   ||||||::::::            |||
     181 laSerMetArgTyrTyrGlnLysPheLeuAspGlnTyrLysProGlnCys 197

    1474 CTTTTCAACAGGCCA 1488
         ::::::|||:::|||
     198 IleLeuAsnLysPro 202

seq_name: SwissProt_39:ATS2_BOVIN

     ocumentation_block:
     ATS2_BOVIN     STANDARD;      PRT;  1205 AA.
     P79331;
     01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ADAM-TS 2 PRECURSOR (EC 3.4.24.14) (A DISINTEGRIN AND
DE   METALLOPROTEINASE WITH THROMBOSPONDIN MOTIFS 2) (ADAMTS-2) (ADAM-TS2)
DE   (PROCOLLAGEN I/II AMINO-PROPEPTIDE PROCESSING ENZYME) (PROCOLLAGEN I
DE   N-PROTEINASE) (PC I-NP) (PROCOLLAGEN N-ENDOPEPTIDASE) (PNPI).
GN   ADAMTS2 OR NPI.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Skin;
RX   MEDLINE=97225960; PubMed=9122202;
RA   Colige A., Li S.W., Sieron A.L., Nusgens B.V., Prockop D.J.,
RA   Lapiere C.M.;
RT   "cDNA cloning and expression of bovine procollagen I N-proteinase: a
RT   new member of the superfamily of zinc-metalloproteinases with binding
RT   sites for cells and other matrix components.";
RL   Proc. Natl. Acad. Sci. U.S.A. 94:2374-2379(1997).
RN   [2]
RP   PARTIAL SEQUENCE.
RX   MEDLINE=95348096; PubMed=7622483;
RA   Colige A., Beschin A., Samyn B., Goebels Y., Van Beeumen J.,
     Nusgens B.V., Lapiere C.M.;
     "Characterization and partial amino acid sequencing of a 107-kDa
     procollagen I N-proteinase purified by affinity chromatography on
     immobilized type XIV collagen.";
     J. Biol. Chem. 270:16724-16730(1995).
CC   -!- FUNCTION: CLEAVES THE PROPEPTIDES OF TYPE I AND II COLLAGEN PRIOR
CC       TO FIBRIL ASSEMBLY. DOES NOT ACT ON TYPE III COLLAGEN. MAY ALSO
CC       PLAY A ROLE IN DEVELOPMENT THAT IS INDEPENDANT OF ITS ROLE IN
CC       COLLAGEN BIOSYNTHESIS.
CC   -!- CATALYTIC ACTIVITY: CLEAVES THE N-PROPEPTIDE OF COLLAGEN CHAIN
CC       ALPHA-1(I) AT PRO-|-GLN AND OF ALPHA-1(II) AND ALPHA-2(I) CHAINS
CC       AT ALA-|-GLN.
CC   -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC   -!- SUBUNIT: MAY BELONG TO A MULTIMERIC COMPLEX. BINDS SPECIFICALLY TO
CC       COLLAGEN TYPE XIV.
CC   -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: ENZYMATIC ACTIVITY IS DETECTED AT HIGH LEVEL
CC       IN ALL TYPE I COLLAGEN-RICH TISSUES SUCH AS SKIN, BONES, TENDONS
CC       AND AORTA AND AT LOW LEVEL IN BRAIN AND THYMUS. THE MRNA LEVELS
CC       WERE DISPROPORTIONALY HIGH IN HEART, LIVER, RETINA AND MUSCLE.
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC   -!- PTM: THE N-TERMINUS IS BLOCKED.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC       SIMILARITY).
CC   -!- DISEASE: DEFECTS IN ADAMTS 2 ARE THE CAUSE OF DERMATOSPARAXIS, A
CC       RECESSIVELY INHERITED DISORDER CHARACTERIZED BY SEVERE SKIN
CC       FRAGILITY AND BIOCHEMICALLY BY THE PRESENCE IN SKIN OF PROCOLLAGEN
CC       INCOMPLETELY PROCESSED AT THE AMINO TERMINUS.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 4 TSP TYPE-1 DOMAINS.
CC   -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS3.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X96389; CAA65253.1; -.
DR   InterPro; IPR000884; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00090; tsp_1; 4.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50092; TSP1; 4.
DR   PROSITE; PS00142; ZINC_PROTEASE; FALSE_NEG.
DR   PROSITE; PS00427; DISINTEGRINS; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Repeat; Collagen degradation; Extracellular matrix; Heparin-binding.
FT   SIGNAL       1     28       POTENTIAL.
FT   PROPEP      29    253       BY SIMILARITY.
FT   CHAIN      254   1205       ADAM-TS 2.
FT   METAL      402    402       ZINC (CATALYTIC) (POTENTIAL).
FT   ACT_SITE   403    403       BY SIMILARITY.
FT   METAL      406    406       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL      412    412       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN     474    554       DISINTEGRIN-LIKE.
FT   DOMAIN     555    611       TSP-TYPE 1 1.
FT   DOMAIN     612    716       CYS-RICH.
FT   DOMAIN     717    845       SPACER.
FT   DOMAIN     846    905       TSP-TYPE 1 2.
FT   DOMAIN     906    968       TSP-TYPE 1 3.
FT   DOMAIN     969   1024       TSP-TYPE 1 4.
FT   SITE       685    687       CELL ATTACHMENT SITE (POTENTIAL).
FT   DOMAIN      31     35       POLY-ALA.
FT   DOMAIN     177    180       POLY-GLU.
FT   CARBOHYD   104    104       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   245    245       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   942    942       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   943    943       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   987    987       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1025   1025       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1092   1092       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1139   1139       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1144   1144       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   1205 AA;  133887 MW;  7B5B232A45320371 CRC64;


alignment_scores:
           Quality:  210.50            Length:   1040
             Ratio:  0.492               Gaps:     55
 Percent Similarity:  41.154   Percent Identity: 20.096

alignment_block:
 US-09-634-252A-2 x ATS2_BOVIN   ..

 Align seg 1/1  to: ATS2_BOVIN  from: 1  to: 1205

     100 CCTGCCAGCGCCCCGGCCCGCACGCCGCCCTGCCGCCTGCTTCTCGTCCT 149
         ||||||:::|||   :::|||         |||      ||||||:::||
       4 ProAlaGlyAlaAlaGlyArgLeu...LeuCysProAlaLeuLeuLeuLe 19

     150 TCTCCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCGCGCCTGGGGGG 199
         ||||||||||   ||||||   |||::::::|||      |||   :::|
```

```
 19 uLeuLeuLeu...ProLeuProAlaAspAlaArgLeuAlaAlaAlaAlaA 35

200 CTGCTGCGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAAAAAAATTTG 249
    ||      |||:::::::|||:::   ::::::   ||||||:::   |||
 35 laAspProProGlyGlyProGln...GlyHisGlyAlaGluArgIleLeu 50

250 GGAGTCCTGGCAGATGAAGACAATACATTGCAACAGAATAGCAGCAGTAA 299
    :::|||   :::      |||
 51 AlaValProValArgThrAsp..................................... 57

300 TATCAGTTACAGCAATGCAATGCAGAAAGAAATCACACTGCCTTCAAGAC 349
                                                  :::||||
 58 .....................................AlaGlnGlyArgL 62

350 TCATATATTACATCAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGAC 399
    ||:::   ::::::   :::   :::   :::|||      |||
 62 euValSerHisValValSerAlaAlaThrAlaProAlaGlyValArgThr 78

400 ACAAAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCA 449
       :::|||         |||                    |||:::
 79 ArgArgAlaAlaProAlaGln............IleProGlyLeuSerGl 91

450 GGCAAGC..........................TTCCAGATTGAAGCCT 472
     :::|||                         :::::::::   :::|
 91 yGlySerGluGluAspProGlyGlyArgLeuPheTyrAsnValThrValP 108

473 TCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCT 522
    |||||      :::   |||   |||      |||      |||::::::
108 heGlyArgAspLeuHisLeuArgLeuArgProAsnAlaArgLeuValAla 124

523 ...TCTGATTATGTGGAGATTCACTACGAAAATGGGAAACCACAGTACTC 569
       :::      ||||||   :::   |||:::|||      :::   ::
125 ProGlyAlaThrValGluTrpGlnGlyGluSerGlyAlaThrArgValGl 141

570 TAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGCGTC...A 616
    :               |||   |||   |||::::::   ||:::   :
141 uProLeuLeuGlyThrCysLeuTyrValGlyAspValAlaGlyLeuAlaG 158

617 AAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTT 666
    :::::|||:::||||||||||||:::|||:::||||||   |||::::::
158 luSerSerSerValAlaLeuSerAsnCysAspGlyLeuAlaGlyLeuIle 174

667 GAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAG......CTGGT 710
    :::   :::          ::::::||||||||||||||      |||::
175 ArgMetGluGluGluGluPhePheIleGluProLeuGluLysGlyLeuAl 191

711 TCATGATGAGAAAAGCACAGGTCGACCACATATAATC..............  747
    :      |||   :::   ||||||   |||::::::
191 aAlaLysGluAlaGluGlnGlyArgValHisValValTyrHisArgProT 208

748 ...............................CAGAAAACCTTGGCAGGA 765
                                   |||        :::|||
208 hrThrSerArgProProProLeuGlyGlyProGlnAlaLeuAspThrGly 224

766 CAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAGGTGACCAGTG 815
        :::   ::::::   :::|||:::         |||
225 IleSerAlaAspSerLeuAspSerLeuSerArgAlaLeuGly........ 238

816 GCCCTTTCTCTCTGAA...TTACAGTGGTTGAAAAGAAGGAAGAGAGCAG 862
          |||:::|||   ::::::      :::||||||   |||
239 ....ValLeuGluGluArgValAsnSerSerArgArgArgMetArg.... 252

863 TGAATCCATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTTATG 912
              |||         ::::::         :::|||::::::
253 ...........ArgHisAlaAlaAspAspAspTyrAsnIleGluValLeu 265

913 ATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACA 962
    :::      |||   ::::::   :::   |||   ::::::      ||
266 LeuGlyValAspAspSerValValGlnPheHisGlyThrGlu......Hi 280

963 TACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACA 1012
    |:::::::::::      :::::::::|||:::|||::::::||||||
280 sValGlnLysTyrLeuLeuThrLeuMetAsnIleValAsnGluIleTyrH 297

1013 AGGAGCAG..............CTCAACACCAGGGTTGTCCTGGTGGCT 1047
       ::::::               :::|||:::   :::|||   :::
 297 isAspGluSerLeuGlyAlaHisIleAsnValValLeuValArgIleIle 313

1048 GTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGT 1097
     :::   ::::::   ::::::   ::::::           ||||||
 314 LeuLeuSerTyrGlyLysSerMetSerLeuIleGluIleGlyAsnProSe 330

1098 GCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATTAAGCAGCATG 1147
      |||   |||::::::      ::::::          :::::|||
 330 rGlnSerLeuGluAsnValCysArgTrpAlaTyr...LeuGlnGlnLysP 346

1148 CTGATGCTGTGCAC.....................CTCATCTCGCGG 1173
       |||:::   |||                          :::::::::|||
 346 roAspThrAspHisAspGluTyrHisAspHisAlaIlePheLeuThrArg 362

1174 GTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGA........ 1215
           |||            |||::::::::::   ::::::
 363 GlnAspPhe......GlyProSerGlyMetGlnGlyTyrAlaProValTh 377

1216 .GGTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAAT...GAGTATGGTC 1261
      |||:::|||      :::|||:::     :::|||   |||   |||:
 377 rGlyMetCysHisProValArgSerCysThrLeuAsnHisGluAspGlyP 394

1262 TTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTT 1311
     ::      |||         |||:::::::::::   ::::::   |||
 394 heSerSerAlaPhe......ValValAlaHisGluThrGlyHisValLeu 408

1312 GGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATGTGACTGC...AC 1358
     |||::::::   :::   ::::::::::::           |||
 409 GlyMetGluHisAspGlyGlnGlyAsnArg............CysGlyAs 421

1359 AGAATCCTGGGGTGGCTGCATCATGGAGGAA......ACAGGGGTGTCCC 1402
      |||         |||   ||||||              ::::::   |
 421 pGluValArgLeuGlySerIleMetAlaProLeuValGlnAlaAlaPheH 438

1403 ATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTA 1452
     ||         :::|||:::||||||      |||         :::|||
 438 isArgPheHisTrpSerArgCysSerGlnGlnGluLeuSerArgTyrLeu 454

1453 CAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGA 1502
     :::      :::      ||||||   :::   |||
 455 His......SerTyrAspCysLeuArgAspAspProPheThrHisAspTr 469

1503 GCCCACG.....................................GAAT 1513
      |||:::                                    :::|
 469 pProAlaLeuProGlnLeuProGlyLeuHisTyrSerMetAsnGluGlnC 486

1514 GT............GGAAATGGATACGTGGAAGCT.............. 1536
     ||            |||   ||||||:::   :::
 486 ysArgPheAspPheGlyLeuGlyTyrMetMetCysThrAlaPheArgThr 502

1536 ..................................................  1536

 503 PheAspProCysLysGlnLeuTrpCysSerHisProAspAsnProTyrPh 519

1537 ............................GGGGAGGAGTGTGATTGTG 1555
                                    |||      |||      |
 519 eCysLysThrLysLysGlyProProLeuAspGlyThrMetCysAlaProG 536

1556 GTTTTCATGTG.........GAATGC..................... 1572
     ||   |||           :::|||
 536 lyLysHisCysPheLysGlyHisCysIleTrpLeuThrProAspIleLeu 552

1573 ............................TATGGATTATGCTGTAA 1589
                                   :::|||   |||   ::
 553 LysArgAspGlyAsnTrpGlyAlaTrpSerProPheGlySerCysSerAr 569

1590 GAAATGTTCC..................................... 1599
     :   |||:::
 569 gThrCysGlyThrGlyValLysPheArgThrArgGlnCysAspAsnProH 586
```

```
1600 .....CTCTCCAACGGGGCTCACTGCAGC.............GACGGGCCC 1632
             :::::::|||      ||||||              |||
 586 isProAlaAsnGlyGlyArgThrCysSerGlyLeuAlaTyrAspPheGln 502

1633 TGCTGTAACAATACCTCATGT............................. 1653
        ||||||:::   :::|||
 603 LeuCysAsnSerGlnAspCysProAspAlaLeuAlaAspPheArgGluGl 519

1654 ...................CTTTTTCAGCCACGAGGG.............. 1671
                         |||:::      :::|||
 619 uGlnCysArgGlnTrpAspLeuTyrPheGluHisGlyAspAlaGlnHisH 536

1672 ...........TATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACT 1710
                 :::|||   |||||||||   ::::::|||   :::
 636 isTrpLeuProHisGluHisArgAspAlaLysGluArgCysHisLeu... 551

1711 GAATATTGT............ACTGGAGACTCTGGTCAGTGCCCACCAAA 1748
        ||||||            ||||||:::      :::
 652 ...TyrCysGluSerLysGluThrGlyGluValValSerMetLysArgMe 667

1749 TCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAG......GGCC 1792
     :::|||      ||||||      |||:::   ::::::      :::
 667 tValHis......AspGlyThrArgCysSerTyrLysAspAlaPheSerL 682

1793 GCTGCTACAATGGCGAGTGCAAGACC....................... 1818
        |||   :::|||:::|||:::
 682 euCysValArgGlyAspCysArgLysValGlyCysAspGlyValIleGly 698

1819 .....................................AGAGACAACCA 1829
                                           ::::::::
 699 SerSerLysGlnGluAspLysCysGlyValCysGlyGlyAspAsnSerHi 715

1830 GTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCT 1879
     :|||:::   :::   ||||||   :::   :::      |||:::   |
 715 sCysLysValValLysGlyThrPheSerArgSerProLysLysLeuGlyT 732

1880 ATGAAAAG.......................................... 1887
     ||   |||
 732 yrIleLysMetPheGluIleProAlaGlyAlaArgHisLeuLeuIleGln 748

1887 ..... 1887

 749 GluAlaAspThrThrSerHisHisLeuAlaValLysAsnLeuGluThrGl 765

1888 ..........CTGAATACAGAA........................... 1899
                |||||   |||
 765 yLysPheIleLeuAsnGluGluAsnAspValAspProAsnSerLysThrP 782

1900 ...........GGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGG 1938
                 |||:::|||                :::|||||   :::
 782 heIleAlaMetGlyValGluTrpGluTyrArgAspGluAspGlyArgGlu 798

1939 TGGATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTAC 1988
        :::|||
 799 ThrLeuGln....................................... 801

1989 CAATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTC 2038
           |||   :::|||      |||   :::          :::||||
 802 .......ThrMetGlyProLeuHisGlyThrIleThrValLeuValIleP 816

2039 CAACTTCC.............................TTCTACCATCAA 2058
     ||   :::                            :::   |||:::
 816 roGluGlyAspAlaArgIleSerLeuThrTyrLysTyrMetIleHisGlu 832

2059 GGCCGGGTG...ATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGA 2105
           :::   :::|||   ::::::          ||||||:::|||||
 833 AspSerLeuAsnValAspAspAsnAsn.........ValLeuGluAspAs 846

2106 TACGGATGTGGGCTAT.........GTAGAAGATGGAACGCCATGT.... 2142
     |:::   |||||||||         ::::::      :::||||||
 846 pSer...ValGlyTyrGluTrpAlaLeuLysLysTrpSerProCysSerL 862

2143 ...........GGCCCGTCTATGATG............TGTTTAGATCGG 2169

                    |||   |||::::::            |||      |||
 862 ysProCysGlyGlyGlySerGlnPheThrLysTyrGlyCys......Arg 876

2170 AAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTC 2219
     :::   |||:::   :::   ::::::   :::   |||          ||
 877 ArgArgLeuAspHisLysMetValHisArgGlyPheCysAspSerValSe 893

2220 CAAGGGTAAA.....:.....GTCTGTTCGGGCCATGGGGTGTGTAGTA 2257
     ||||   |||            :::|||:::   :::      ||||||:
 893 rLysProLysAlaIleArgArgThrCysAsnProGln...GluCysSerG 909

2258 ATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGC... 2304
     ::   :::            ::::::      ::::::   :::|||
 909 lnProValTrpValThrGlyGluTrpGluProCysSerArgSerCysGly 925

2305 .....................AGTATCCGG..................... 2313
                           |||:::|||
 926 ArgThrGlyMetGlnValArgSerValArgCysValGlnProLeuHisAs 942

2314 .GATCCAGTTAGGAACCTTCACCCCCCC......AAGGATGAAGGACCCA 2356
      :::   :::|||::::::|||            :::|||      |||:
 942 nAsnThrThrArgSerValHisThrLysHisCysAsnAspAlaArgProG 959

2357 AGGGTCCTAGTGCCACCAAT 2376
     ::|||      |||   |||
 959 luGlyArgArgAlaCysAsn 965
```

seq_name: SwissProt_39:TENA_CHICK

seq_documentation_block:

ID TENA_CHICK STANDARD; PRT; 1808 AA.
AC P10039; P13132; O73584; O73585;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TENASCIN PRECURSOR (TN) (HEXABRACHION) (CYTOTACTIN) (NEURONECTIN)
DE (GMEM) (JI) (MIOTENDINOUS ANTIGEN) (GLIOMA-ASSOCIATED-EXTRACELLULAR
DE MATRIX ANTIGEN) (GP 150-225).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=90030407; PubMed=2478295;
RA Spring J., Beck K., Chiquet-Ehrismann R.;
RT "Two contrary functions of tenascin: dissection of the active sites
RT by recombinant tenascin fragments.";
RL Cell 59:325-334(1989).
RN [2]
RP SEQUENCE OF 27-722 FROM N.A., AND SEQUENCE OF 79-96.
RC TISSUE=Fibroblast;
RX MEDLINE=89030589; PubMed=2460335;
RA Pearson C.A., Pearson D., Shibahara S., Hofsteenge J.,
RA Chiquet-Ehrismann R.;
RT "Tenascin: cDNA cloning and induction by TGF-beta.";
RL EMBO J. 7:2977-2982(1988).
RN [3]
RP SEQUENCE OF 464-1018 AND 1412-1661 FROM N.A., AND SEQUENCE OF 852-868.
RC TISSUE=Embryo;
RX MEDLINE=88176910; PubMed=2451243;
RA Jones F.S., Burgoon M.P., Hoffman S., Crossin K.L., Cunningham B.A.,
RA Edelman G.M.;
RT "A cDNA clone for cytotactin contains sequences similar to epidermal
RT growth factor-like repeats and segments of fibronectin and
RT fibrinogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:2186-2190(1988).
CC -!- FUNCTION: SAM (SUBSTRATE-ADHESION MOLECULE) THAT APPEARS TO
CC INHIBIT CELL MIGRATION. MAY PLAY A ROLE IN SUPPORTING THE GROWTH
CC OF EPITHELIAL TUMORS.
CC -!- SUBUNIT: HEXAMERIC. A HOMOTRIMER MAY BE FORMED IN THE TRIPLE

```
CC       COILED-COIL REGION AND MAY BE STABILIZED BY DISULFIDE RINGS AT
CC       BOTH ENDS. TWO OF SUCH HALF-HEXABRACHIONS MAY BE DISULFIDE LINKED
CC       WITHIN THE CENTRAL GLOBULE.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 230 KDA (SHOWN HERE), 200 KDA
CC       AND 190 KDA; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY ARE
CC       PRODUCED IN A TISSUE-AND TIME-SPECIFIC MANNER DURING DEVELOPMENT.
CC   -!- INDUCTION: BY TGF-BETA.
CC   -!- SIMILARITY: CONTAINS 13.5 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 11 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS A DOMAIN RELATED TO THE C-TERMINAL PART OF
CC       THE BETA AND GAMMA CHAINS OF FIBRINOGEN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M23121; AAA49086.1; -.
DR   EMBL; X08031; CAB40811.1; -.
DR   EMBL; X08030; CAA30824.1; ALT_TERM.
DR   EMBL; J03641; AAA48748.1; ALT_SEQ.
DR   EMBL; M20816; AAA48749.1; ALT_SEQ.
DR   PIR; A30903; A30903.
DR   PIR; A31930; A31930.
DR   PIR; A33379; A33379.
DR   PIR; B33379; B33379.
DR   PIR; C33379; C33379.
DR   PIR; S01292; S01292.
DR   HSSP; P24821; 1TEN.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001777; -.
DR   InterPro; IPR002181; -.
DR   Pfam; PF00008; EGF; 13.
DR   Pfam; PF00147; fibrinogen_C; 1.
DR   Pfam; PF00041; fn3; 11.
DR   PROSITE; PS00022; EGF_1; 14.
DR   PROSITE; PS01186; EGF_2; 14.
KW   Glycoprotein; Cell adhesion; Repeat; EGF-like domain; Coiled coil;
KW   Extracellular matrix; Signal; Alternative splicing.
FT   SIGNAL       1     22
FT   PROPEP      23     33
FT   CHAIN        34   1808       TENASCIN.
FT   DOMAIN      118    142       COILED COIL (POTENTIAL).
FT   DOMAIN      176    188       EGF-LIKE 1 (INCOMPLETE).
FT   DOMAIN      188    219       EGF-LIKE 2.
FT   DOMAIN      219    250       EGF-LIKE 3.
FT   DOMAIN      250    281       EGF-LIKE 4.
FT   DOMAIN      281    312       EGF-LIKE 5.
FT   DOMAIN      312    343       EGF-LIKE 6.
FT   DOMAIN      343    374       EGF-LIKE 7.
FT   DOMAIN      374    405       EGF-LIKE 8.
FT   DOMAIN      405    436       EGF-LIKE 9.
FT   DOMAIN      436    467       EGF-LIKE 10.
FT   DOMAIN      467    498       EGF-LIKE 11.
FT   DOMAIN      498    529       EGF-LIKE 12.
FT   DOMAIN      529    560       EGF-LIKE 13.
FT   DOMAIN      560    591       EGF-LIKE 14.
FT   DOMAIN      591    680       FIBRONECTIN TYPE-III 1.
FT   DOMAIN      681    771       FIBRONECTIN TYPE-III 2.
FT   DOMAIN      772    862       FIBRONECTIN TYPE-III 3.
FT   DOMAIN      863    954       FIBRONECTIN TYPE-III 4.
FT   DOMAIN      955   1042       FIBRONECTIN TYPE-III 5.
FT   DOMAIN     1043   1133       FIBRONECTIN TYPE-III 6.
FT   DOMAIN     1134   1224       FIBRONECTIN TYPE-III 7.
FT   DOMAIN     1225   1315       FIBRONECTIN TYPE-III 8.
FT   DOMAIN     1316   1404       FIBRONECTIN TYPE-III 9.
FT   DOMAIN     1405   1492       FIBRONECTIN TYPE-III 10.
FT   DOMAIN     1493   1580       FIBRONECTIN TYPE-III 11.
FT   DOMAIN     1589   1808       FIBRINOGEN BETA/GAMMA.
FT   DISULFID     64     64       INTERCHAIN (POTENTIAL).
FT   DISULFID    192    202       BY SIMILARITY.
FT   DISULFID    196    207       BY SIMILARITY.
FT   DISULFID    209    218       BY SIMILARITY.
FT   DISULFID    223    233       BY SIMILARITY.
FT   DISULFID    227    238       BY SIMILARITY.
FT   DISULFID    240    249       BY SIMILARITY.
FT   DISULFID    254    264       BY SIMILARITY.
FT   DISULFID    258    269       BY SIMILARITY.
FT   DISULFID    271    280       BY SIMILARITY.
FT   DISULFID    285    295       BY SIMILARITY.
FT   DISULFID    289    300       BY SIMILARITY.
FT   DISULFID    302    311       BY SIMILARITY.
FT   DISULFID    316    326       BY SIMILARITY.
FT   DISULFID    320    331       BY SIMILARITY.
FT   DISULFID    333    342       BY SIMILARITY.
FT   DISULFID    347    357       BY SIMILARITY.
FT   DISULFID    351    362       BY SIMILARITY.
FT   DISULFID    364    373       BY SIMILARITY.
FT   DISULFID    378    388       BY SIMILARITY.
FT   DISULFID    382    393       BY SIMILARITY.
FT   DISULFID    395    404       BY SIMILARITY.
FT   DISULFID    409    419       BY SIMILARITY.
FT   DISULFID    413    424       BY SIMILARITY.
FT   DISULFID    426    435       BY SIMILARITY.
FT   DISULFID    440    450       BY SIMILARITY.
FT   DISULFID    444    455       BY SIMILARITY.
FT   DISULFID    457    466       BY SIMILARITY.
FT   DISULFID    471    481       BY SIMILARITY.
FT   DISULFID    475    486       BY SIMILARITY.
FT   DISULFID    488    497       BY SIMILARITY.
FT   DISULFID    502    512       BY SIMILARITY.
FT   DISULFID    506    517       BY SIMILARITY.
FT   DISULFID    519    528       BY SIMILARITY.
FT   DISULFID    533    543       BY SIMILARITY.
FT   DISULFID    537    548       BY SIMILARITY.
FT   DISULFID    550    559       BY SIMILARITY.
FT   DISULFID    564    574       BY SIMILARITY.
FT   DISULFID    568    579       BY SIMILARITY.
FT   DISULFID    581    590       BY SIMILARITY.
FT   CARBOHYD     38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    168    168       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    186    186       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    328    328       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    603    603       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    643    643       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    751    751       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    759    759       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1050   1050       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1090   1090       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1101   1101       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1112   1112       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1153   1153       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1183   1183       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1416   1416       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1736   1736       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1769   1769       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   1043   1224       MISSING (IN 200 KDA ISOFORM).
FT   VARSPLIC   1043   1315       MISSING (IN 190 KDA ISOFORM).
FT   CONFLICT    182    182       W -> R (IN REF. 2).
FT   CONFLICT    563    571       SCPNDCNNV -> PAPMTATTW (IN REF. 3).
FT   CONFLICT    598    598       E -> G (IN REF. 3).
FT   CONFLICT    838    838       T -> TEY (IN REF. 3).
FT   CONFLICT    886    886       N -> F (IN REF. 3).
SQ   SEQUENCE   1808 AA;  198858 MW;  B924A06CF9EFD6DE CRC64;


alignment_scores:
           Quality:  209.50            Length:    498
             Ratio:  1.017               Gaps:     42
 Percent Similarity:  41.365   Percent Identity: 26.104

alignment_block:
```

```
US-09-634-252A-2 x TENA_CHICK   ..

Align seq 1/1  to: TENA_CHICK  from: 1  to: 1808

  1143 GCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGA 1192
       ||||||   |||:::|||         |||   ||||||
   110 AlaCysGlyCysAlaAla.........AlaProAspIle........... 119

  1193 GAAGCAGTCTGAGTTACTTTGGAGGTGT..............CTGTTCT 1227
              ::::::||||||   |||                  |||   |
   120 .......LysAspLeuLeuSerArgLeuGluGluLeuGluGlyLeuValS 134

  1228 CGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTCGC 1277
       ||      |||:::   |||               ::::::|||:::|||
   134 erSerLeuArgGluGlnCys..............AlaSerGlyAlaGly 145

  1277 ..................................................... 1277

   146 CysCysProAsnSerGlnThrAlaGluGlyArgLeuAspThrAlaProTy 162

  1278 .........................ACAAGTATT.........ATCGCAG 1293
                                |||:::|||         ::::::|
   162 rCysSerGlyHisGlyAsnTyrSerThrGluIleCysGlyCysValCysG 179

  1294 AGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCC 1343
       ||||||||   |||               ||| |||:::      :::||
   179 luProGlyTrpLys..............Gly.ProAsnCysSerGluPr 190

  1344 AAAATGT......GACTGCACAGAATCCTGG............GGTGGCT 1375
       |   |||      :::|||   :::                   |||   |
   190 oAlaCysProArgAsnCysLeuAsnArgGlyLeuCysValArgGlyLysC 207

  1376 GCATCATGGAGGAA.....ACAGGGGTGTCCCATTCTCGAAAATTT... 1416
       |||||   ||||||      ||||||   :::   |||:::
   207 ysIleCysGluGluGlyPheThrGlyGluAspCysSerGlnAlaAlaCys 223

  1417 ...TCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGG 1463
          |||   |||:::              |||      :::
   224 ProSerAspCysAsn..............AspGlnGlyLysCysValAs 235

  1464 TGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAAT 1513
        |||:::|||:::                                    |
   235 pGlyValCysVal.....................................C 240

  1514 GTGGAAATGGATACGTG......GAAGCTGGGGAGGAG...TGT...... 1548
       ||   :::||||||:::      ::::::|||||||||   |||
   240 ysPheGluGlyTyrThrGlyProAspCysGlyGluGluLeuCysProHis 256

  1549 GATTGTGGTTTTCATGTGGAATGCTATGGATTATGCTGTAAGAAATGTTC 1598
          ||||||:::|||   :::|||   |||      |||       ::
   257 GlyCysGlyIleHisGlyArgCysValGlyGlyArgCysValCysHisGl 273

  1599 CCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCT 1648
       :      :::|||      |||::::::      |||   ||||||
   273 uGlyPheThrGlyGluAspCysAsnGluProLeuCysProAsnAsn.... 288

  1649 CATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAG 1698
         |||      :::   ||||||   :::|||   |||       ||||||
   289 ..Cys......HisAsnArgGly...ArgCysValAsp......AsnGlu 299

  1699 TGTGATATTACTGAATATTGTACTGGA...GACTCTGGTCAG......TG 1739
       |||         |||      ||||||   |||   |||:::      ||
   300 CysValCysAspGluGlyTyrThrGlyGluAspCysGlyGluLeuIleCy 316

  1740 CCCACCAAATCTTCATAAG............................... 1758
       ||||   :::
   316 sProAsnAspCysPheAspArgGlyArgCysIleAsnGlyThrCysPheC 333

  1759 ..CAAGACGGATATGCA................................. 1773
         ::::::||||||:::
   333 ysGluGluGlyTyrThrGlyGluAspCysGlyGluLeuThrCysProAsn 349

  1774 ...TGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAG 1820

         ||||||   |||   |||||||||   ||||||   |||      ::
   350 AsnCysAsnGlyAsn...GlyArgCysGluAsnGlyLeuCysValCysHi 365

  1821 AGAC............AACCAGTGTCAGTACATCTGGGGAACAAAGGCTG 1858
       ::::                ::::::|||
   365 sGluGlyPheValGlyAspAspCys........................ 373

  1859 CAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAG 1908
            |||:::|||   |||   :::      |||             :::
   374 .....SerGlnLysArgCysProLysAspCysAsn............Asn 384

  1909 AAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACA 1958
       :::|||:::|||      ||||||                :::|||      ||
   385 ArgGlyHisCys...ValAspGly............ArgCysValCysHi 396

  1959 TGATGTGTTC............TGTGGATTCTTACTCTGTACCAATCTTA 1996
       |:::   :::             ||||||   |||   |||
   396 sGluGlyTyrLeuGlyGluAspCysGlyGluLeuArgCys.......... 409

  1997 CTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCC 2046
                                                  |||::::::
   410 ......................................ProAsnAsp 412

  2047 TTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTT 2096
          :::::::::||||||         |||   :::::::::   |||
   413 CysHisAsnArgGlyArg.........CysIleAsnGlyGlnCysValCy 426

  2097 AGATGATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCC 2146
        |||:::            |||::::::      |||      ||||||
   426 sAspGlu............GlyPheIle......GlyGluAspCysGlyG 437

  2147 CGTCTATGATG.........TGTTTAGATCGG...AAGTGCCTACAAATT 2184
                           |||   :::|||   :::|||::::::
   437 luLeuArgCysProAsnAspCysHisAsnArgGlyArgCysValAsnGly 453

  2185 CAAGCC............................CTAAATATGAGCAG 2204
       |||:::                          :::
   454 GlnCysGluCysHisGluGlyPheIleGlyGluAspCysGlyGluLeuAr 470

  2205 CTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTA 2254
        ||||||   |||               |||::::::||||||   |||
   470 gCysProAsnAsp..............CysAsnSerHisGlyArgCysV 482

  2255 GTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGC 2304
         |||   :::   |||:::||||||      ::::::|||   ||||||
   482 alAsn...GlyGlnCysValCysAspGluGlyTyrThrGlyGluAspCys 497

  2305 AGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAAGGAT 2346
       :::                  :::|||:::   |||:::|||
   498 Gly.................GluLeuArgCysProAsnAsp 505

seq_name: SwissProt_39:TENA_PIG

seq_documentation_block:
ID   TENA_PIG       STANDARD;      PRT;  1746 AA.
AC   Q29116; P98142;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   TENASCIN PRECURSOR (TN) (HEXABRACHION) (CYTOTACTIN) (NEURONECTIN)
DE   (GMEM) (JI) (MIOTENDINOUS ANTIGEN) (GLIOMA-ASSOCIATED-EXTRACELLULAR
DE   MATRIX ANTIGEN) (GP 150-225) (TENASCIN-C) (TN-C) (P230).
GN   HXB.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Submaxillary gland;
RX   MEDLINE=92104189; PubMed=1722152;
RA   Nishi T., Weinstein J., Gillespie W.M., Paulson J.C.;
RT   "Complete primary structure of porcine tenascin: detection of
```

```
RT   tenascin transcript in adult submaxillary glands.";
RL   Eur. J. Biochem. 202:643-648(1991).
RN   [2]
RP   SEQUENCE OF 813-825; 887-917; 998-1011; 1597-1608 AND 1719-1730.
RC   TISSUE=Fetal brain;
RX   MEDLINE=98158323; PubMed=9498558;
RA   Wakatsuki S., Ho S.H., Arioka M., Yamasaki M., Kitamoto K.;
RT   "Isolation and characterization of a 230 kDa protein (p230)
RT   specifically expressed in fetal brains: its involvement in neurite
RT   outgrowth from rat cerebral cortex neurons grown on monolayer of
RT   astrocytes.";
RL   J. Biochem. 122:1146-1152(1997).
CC   -!- FUNCTION: SAM (SUBSTRATE-ADHESION MOLECULE) THAT APPEARS TO
CC       INHIBIT CELL MIGRATION. MAY PLAY A ROLE IN SUPPORTING THE GROWTH
CC       OF EPITHELIAL TUMORS.
CC   -!- FUNCTION: PLAYS A ROLE DURING EARLY BRAIN DEVELOPMENT PARTICULARLY
CC       IN GROWTH CONE GUIDANCE. INVOLVED IN NEURITE OUTGROWTH FROM
CC       CORTICAL NEURONS GROWN ON THE MONOLAYER OF ASTROCYTES.
CC   -!- SUBUNIT: HEXAMERIC. A HOMOTRIMER MAY BE FORMED IN THE TRIPLE
CC       COILED-COIL REGION AND MAY BE STABILIZED BY DISULFIDE RINGS AT
CC       BOTH ENDS. TWO OF SUCH HALF-HEXABRACHIONS MAY BE DISULFIDE LINKED
CC       WITHIN THE CENTRAL GLOBULE.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; MAJOR, MINOR-1 AND MINOR-2
CC       (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING. THEY ARE
CC       PRODUCED IN A TISSUE-AND TIME-SPECIFIC MANNER DURING DEVELOPMENT.
CC   -!- TISSUE SPECIFICITY: SUBMAXILLARY GLANDS AND BRAIN.
CC   -!- DEVELOPMENTAL STAGE: PREDOMINANTLY EXPRESSED IN THE EMBRYONIC AND
CC       EARLY POSTNATAL STAGES. LITTLE OR NO DETECTION IN ADULT BRAIN.
CC   -!- INDUCTION: BY TGF-BETA.
CC   -!- SIMILARITY: CONTAINS 15 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 11 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 FIBRINOGEN-LIKE DOMAIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X61599; CAA43796.1; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001777; -.
DR   InterPro; IPR002049; -.
DR   InterPro; IPR002181; -.
DR   Pfam; PF00008; EGF; 14.
DR   Pfam; PF00147; fibrinogen_C; 1.
DR   Pfam; PF00041; fn3; 10.
DR   PRINTS; PR00011; EGFLAMININ.
DR   PRINTS; PR00014; FNTYPEIII.
DR   PROSITE; PS00022; EGF_1; 15.
DR   PROSITE; PS01186; EGF_2; 14.
KW   Glycoprotein; Cell adhesion; Repeat; EGF-like domain; Coiled coil;
KW   Extracellular matrix; Alternative splicing; Signal.
FT   SIGNAL       1     22       POTENTIAL.
FT   CHAIN       23   1746       TENASCIN.
FT   DOMAIN      23    185       INVOLVED IN HEXAMER FORMATION.
FT   DOMAIN     118    145       COILED COIL (POTENTIAL).
FT   DOMAIN     174    186       EGF-LIKE 1 (INCOMPLETE).
FT   DOMAIN     187    217       EGF-LIKE 2.
FT   DOMAIN     218    249       EGF-LIKE 3.
FT   DOMAIN     250    280       EGF-LIKE 4.
FT   DOMAIN     281    311       EGF-LIKE 5.
FT   DOMAIN     312    342       EGF-LIKE 6.
FT   DOMAIN     343    373       EGF-LIKE 7.
FT   DOMAIN     374    404       EGF-LIKE 8.
FT   DOMAIN     405    435       EGF-LIKE 9.
FT   DOMAIN     436    466       EGF-LIKE 10.
FT   DOMAIN     467    497       EGF-LIKE 11.
FT   DOMAIN     498    528       EGF-LIKE 12.
FT   DOMAIN     529    559       EGF-LIKE 13.
FT   DOMAIN     560    589       EGF-LIKE 14.
FT   DOMAIN     590    620       EGF-LIKE 15.
FT   DOMAIN     621    710       FIBRONECTIN TYPE-III 1.
FT   DOMAIN     711    801       FIBRONECTIN TYPE-III 2.
FT   DOMAIN     802    891       FIBRONECTIN TYPE-III 3.
FT   DOMAIN     892    983       FIBRONECTIN TYPE-III 4.
FT   DOMAIN     984   1071       FIBRONECTIN TYPE-III 5.
FT   DOMAIN    1072   1162       FIBRONECTIN TYPE-III 6.
FT   DOMAIN    1163   1253       FIBRONECTIN TYPE-III 7.
FT   DOMAIN    1254   1342       FIBRONECTIN TYPE-III 8.
FT   DOMAIN    1343   1430       FIBRONECTIN TYPE-III 9.
FT   DOMAIN    1431   1518       FIBRONECTIN TYPE-III 10.
FT   DOMAIN    1527   1733       FIBRINOGEN BETA/GAMMA.
FT   DISULFID    64     64       INTERCHAIN (POTENTIAL).
FT   DISULFID   190    200       BY SIMILARITY.
FT   DISULFID   194    205       BY SIMILARITY.
FT   DISULFID   207    216       BY SIMILARITY.
FT   DISULFID   221    231       BY SIMILARITY.
FT   DISULFID   225    236       BY SIMILARITY.
FT   DISULFID   238    247       BY SIMILARITY.
FT   DISULFID   252    263       BY SIMILARITY.
FT   DISULFID   256    268       BY SIMILARITY.
FT   DISULFID   270    279       BY SIMILARITY.
FT   DISULFID   284    294       BY SIMILARITY.
FT   DISULFID   288    299       BY SIMILARITY.
FT   DISULFID   301    310       BY SIMILARITY.
FT   DISULFID   315    325       BY SIMILARITY.
FT   DISULFID   319    330       BY SIMILARITY.
FT   DISULFID   332    341       BY SIMILARITY.
FT   DISULFID   346    356       BY SIMILARITY.
FT   DISULFID   350    361       BY SIMILARITY.
FT   DISULFID   363    372       BY SIMILARITY.
FT   DISULFID   377    387       BY SIMILARITY.
FT   DISULFID   381    392       BY SIMILARITY.
FT   DISULFID   394    403       BY SIMILARITY.
FT   DISULFID   408    418       BY SIMILARITY.
FT   DISULFID   412    423       BY SIMILARITY.
FT   DISULFID   425    434       BY SIMILARITY.
FT   DISULFID   439    449       BY SIMILARITY.
FT   DISULFID   443    454       BY SIMILARITY.
FT   DISULFID   456    465       BY SIMILARITY.
FT   DISULFID   470    480       BY SIMILARITY.
FT   DISULFID   474    485       BY SIMILARITY.
FT   DISULFID   487    496       BY SIMILARITY.
FT   DISULFID   501    511       BY SIMILARITY.
FT   DISULFID   505    516       BY SIMILARITY.
FT   DISULFID   518    527       BY SIMILARITY.
FT   DISULFID   532    542       BY SIMILARITY.
FT   DISULFID   536    547       BY SIMILARITY.
FT   DISULFID   549    558       BY SIMILARITY.
FT   DISULFID   563    573       BY SIMILARITY.
FT   DISULFID   567    578       BY SIMILARITY.
FT   DISULFID   580    589       BY SIMILARITY.
FT   DISULFID   594    604       BY SIMILARITY.
FT   DISULFID   598    609       BY SIMILARITY.
FT   DISULFID   611    620       BY SIMILARITY.
FT   CARBOHYD    38     38       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   166    166       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   184    184       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   327    327       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   788    788       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1034   1034       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1079   1079       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1121   1121       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD  1354   1354       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC  1072   1162       MISSING (IN MAJOR ISOFORM).
FT   VARSPLIC  1072   1253       MISSING (IN MINOR-1 ISOFORM).
FT   CONFLICT  1007   1007       T -> M (IN REF. 2).
SQ   SEQUENCE  1746 AA;  191399 MW;  56549B1CFE5E5C88 CRC64;


alignment_scores:
                Quality:  209.00            Length:    510
```

```
            Ratio:  1.020                Gaps:     36
 Percent Similarity:  40.196   Percent Identity: 23.922

alignment_block:
 US-09-634-252A-2 x TENA_PIG   ..

 Align seg 1/1  to: TENA_PIG  from: 1  to: 1746

   1073 AGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCT..... 1117
        |||:::::::   |||         |||   |||:::   :::
    103 ArgIleAsnIleProArgArgAlaCysGlyCysAlaAlaAlaProAspVa 119

   1118 .....CAAAATACCGGCAGCGCATTAAGCAGCA 1145
                        ::::::      :::   :::||||||:
    119 lLysGluLeuLeuSerArgLeuGluGluLeuGluAsnLeuValSerSerL 136

   1146 TGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAA 1195
        ::         |||||||||::: ||||||
    136 euArgGluGlnCysThrSerGly.AlaGly..................... 145

   1196 GCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCG.............. 1229
                          |||   ::::::
    146 .....CysCysLeuGlnProAlaGluGlyArgL 155

   1230 ..CACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGC 1277
              |||   :::|||:::          :::   :::   :::|||
    155 euAspThrArgProPheCysSerGlyArgGlyAsnPheSerThrGluGly 171

   1278 ACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAAC 1327
           :::   ::::::|||||||||   |||                ||| |
    172 CysGlyCysValCysGluProGlyTrpLys.............GlyAP 183

   1328 CTTCTAGCAGAAAGCCAAAATGT.....GAC 1353
        ||:::      :::|||:::|||        :::
    183 roAsnCysSerGluProGluCysProSerAsnCysHisLeuArgGlyGln 199

   1354 TGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAA......ACAGGGGT 1397
        |||::::::      |||   |||:::   :::|||      ||||||
    200 CysValAsp......GlyGlnCysValCysAsnGluGlyPheThrGly.. 213

   1398 GTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACT 1447
                        :::   ||||||   |||
    214 ..................GluAspCysSerGlnLeu..... 219

   1448 TTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTA 1497
                         ||||||            |||:::
    220 ..................AlaCys...........ProSerAspCys 225

   1498 TTTGAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTG 1547
           :::      :::|||   ||||||                    ||
    226 AsnAspGlnGlyLysCysValAsnGlyVal.................Cy 236

   1548 TGATTGT......GGTTTTCAT...GTGGAATGCTATGGATTATGCTGTA 1588
        |   |||      |||:::     |||:::|||          |||
    236 sValCysPheGluGlyTyrSerGlyValAspCysSerArgGluThrCysP 253

   1589 AG...AAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGC 1635
             ||||||   ::::::|||   :::|||   ||||||   |||
    253 roValProCysSerGluGluHisGly...ArgCysValAspGlyArgCys 268

   1636 ............................TGTAACAATACCTCATGTCT 1655
                                    ||||||:::      |||||
    269 ValCysGlnGluGlyPheAlaGlyGluAspCysAsnGluProLeuCysLe 285

   1656 TTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATA 1705
        |   :::   :::|||   :::   |||   :::   |||||||||
    285 uHisAsnCysHisGly...ArgGlyArgCysValGluAsnGluCysValC 301

   1706 TTACTGAATATTGTACTGGA...GACTCTGGTCAG......TGCCCACCA 1746
             |||      ||||||   |||   |||:::      ||||||
    301 ysAspGluGlyPheThrGlyGluAspCysGlyGluLeuIleCysProLys 317

   1747 AATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTG 1796
```

```
        :::                                  ::::::||||||||
    318 AspCysPhe.........................AspArgGlyArgCy 325

   1797 CTACAATGGCGAGTGC................................ 1812
        |   ||||||   |||
    325 sIleAsnGlyThrCysTyrCysAspGluGlyPheGluGlyGluAspCysG 342

   1813 .....................................AAGACCAGAGAC 1824
                                             :::   ::::::
    342 lyArgLeuAlaCysProHisGlyCysArgGlyArgGlyArgCysGluGlu 358

   1825 AACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTT 1874
        :::||||||      :::                 ||||||:::|||
    359 GlyGlnCys......ValCysAspGluGlyPheAlaGlyAlaAsp..... 371

   1875 CTGCTATGAA...AAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCG 1921
         |||   |||   :::      ::::::     ::::::|||:::|||
    372 .CysSerGluArgArgCysProSerAspCysHisAsnArgGlyArgCys. 387

   1922 GGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTC... 1968
              ||||||            :::|||:::      |||   |||
    388 ..LeuAspGly............ArgCysGluCysAspAspGlyPheGlu 399

   1969 .........TGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACG 2009
                   ||||||   |||   |||
    400 GlyGluAspCysGlyGluLeuArgCys...................... 408

   2010 TATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAG 2059
                                    |||   :::          :::|
    409 ............................ProGlyGlyCysSerGlyHisG 416

   2060 GCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGAT...... 2103
        |||||          |||   :::::::::   |||   |||:::
    416 lyArg.........CysValAsnGlyGlnCysValCysAspGluGlyArg 429

   2104 ...GATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTC 2150
                  |||   :::   ::::::      :::   |||
    430 ThrGlyGluAspCysSerGlnLeuArgCysProAsnAspCysHisGlyAr 446

   2151 TATGATGTGTTTAGATCGGAAGTGC..............CTACAAATTC 2185
              |||::::::   :::|||               :::|||
    446 gGlyArgCysValGlnGlyArgCysGluCysGluHisGlyPheGlnGlyT 463

   2186 AAGCCCTAAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGT 2235
              :::      |||||||||   |||                |||
    463 yrAspCysSerGluMetSerCysProHisAsp..............Cys 474

   2236 TCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCAC 2285
              ||||||   |||   |||   :::   |||:::||||||
    475 HisGlnHisGlyArgCysValAsn...GlyMetCysValCysAspAspGl 490

   2286 CTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACC 2335
         ::::::|||   ||||||                |||:::|||:::
    490 yTyrThrGlyGluAspCys.................ArgGluLeuArgC 501

   2336 CCCCCAAGGATGAAGGACCCAAGGGT 2361
          |||   |||   :::   :::|||
    501 ysProGlyAspCysSerGlnArgGly 509

seq_name: SwissProt_39:ATS4_HUMAN

seq_documentation_block:
ID   ATS4_HUMAN     STANDARD;      PRT;   837 AA.
AC   O75173; Q9UN83;
DT   01-OCT-2000 (Rel. 40, Created)
DT   01-OCT-2000 (Rel. 40, Last sequence update)
DT   01-OCT-2000 (Rel. 40, Last annotation update)
DE   ADAM-TS 4 PRECURSOR (EC 3.4.24.-) (A DISINTEGRIN AND METALLOPROTEINASE
DE   WITH THROMBOSPONDIN MOTIFS 4) (ADAMTS-4) (ADAM-TS4) (AGGRECANASE 1)
DE   (ADMP-1).
GN   ADAMTS4 OR KIAA0688.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98403880; PubMed=9734811;
RA   Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA   Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. X.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:169-176(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99286303; PubMed=10356395;
RA   Tortorella M.D., Burn T.C., Pratta M.A., Abbaszade I., Hollis J.M.,
RA   Liu R., Rosenfeld S.A., Copeland R.A., Decicco C.P., Wynn R.,
RA   Rockwell A., Yang F., Duke J.L., Solomon K., George H., Bruckner R.,
RA   Nagase H., Itoh Y., Ellis D.M., Ross H., Wiswall B.H., Murphy K.,
RA   Hillman M.C. Jr., Hollis G.F., Newton R.C., Magolda R.L.,
     Trzaskos J.M., Arner E.C.;
     "Purification and cloning of aggrecanase-1: a member of the ADAMTS
     family of proteins.";
     Science 284:1664-1666(1999).
     [3]
RP   PARTIAL SEQUENCE, AND CHARACTERIZATION.
RX   MEDLINE=20400518; PubMed=10827174;
RA   Tortorella M., Pratta M., Liu R.Q., Abbaszade I., Ross H., Burn T.,
RA   Arner E.;
RT   "The thrombospondin motif of aggrecanase-1 (ADAMTS-4) is critical for
RT   aggrecan substrate recognition and cleavage.";
RL   J. Biol. Chem. 275:25791-25797(2000).
CC   -!- FUNCTION: CLEAVES AGGRECAN, A CARTILAGE PROTEOGLYCAN, AND MAY BE
CC       INVOLVED IN ITS TURNOVER. MAY PLAY AN IMPORTANT ROLE IN THE
CC       DESTRUCTION OF AGGRECAN IN ARTHRITIC DISEASES.
CC   -!- CATALYTIC ACTIVITY: CLEAVES AGGRECAN AT THE 392-GLU-|-ALA-393
CC       SITE.
CC   -!- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: SECRETED. ASSOCIATED WITH THE EXTRACELLULAR
CC       MATRIX (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LUNG AND HEART. EXPRESSED
CC       AT VERY LOW LEVEL IN PLACENTA AND SKELETAL MUSCLES.
CC   -!- INDUCTION: BY INTERLEUKIN 1.
CC   -!- DOMAIN: THE SPACER DOMAIN AND THE TSP TYPE 1 DOMAINS ARE IMPORTANT
CC       FOR A TIGHT INTERACTION WITH THE EXTRACELLULAR MATRIX.
CC   -!- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 DISINTEGRIN-LIKE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 TSP TYPE-1 DOMAIN.
CC   -!- CAUTION: HAS SOMETIMES BEEN REFERRED TO ADAMTS2.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB014588; BAA31663.1; -.
DR   EMBL; AF148213; AAD41494.1; -.
DR   MIM; 603876; -.
DR   HSSP; P34179; 1IAG.
DR   MEROPS; M12.221; -.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR000884; -.
DR   InterPro; IPR001590; -.
DR   Pfam; PF00090; tsp_1; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50092; TSP1; 1.
DR   PROSITE; PS00427; DISINTEGRINS; FALSE_NEG.
KW   Hydrolase; Metalloprotease; Zinc; Signal; Glycoprotein; Zymogen;
KW   Extracellular matrix.
FT   SIGNAL       1     51       POTENTIAL.
FT   PROPEP      52    212
FT   CHAIN      213    837       ADAM-TS 4.
FT   SITE       194    194       CYSTEINE SWITCH (POTENTIAL).
FT   METAL      361    361       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE   362    362       BY SIMILARITY.
FT   METAL      365    365       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL      371    371       ZINC (CATALYTIC) (BY SIMILARITY).
FT   DOMAIN     437    519       DISINTEGRIN-LIKE.
FT   DOMAIN     520    576       TSP-TYPE 1 1.
FT   DOMAIN     577    685       CYS-RICH.
FT   DOMAIN     686    837       SPACER.
FT   DOMAIN     247    252       POLY-ALA.
FT   CARBOHYD    68     68       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT    77     77       A -> T (IN REF. 1).
SQ   SEQUENCE   837 AA;  90224 MW;  5DF9C9AC137DF41F CRC64;


alignment_scores:
            Quality:  208.50              Length:    853
              Ratio:  0.589                 Gaps:     44
 Percent Similarity:  41.501   Percent Identity: 20.750

alignment_block:
 US-09-634-252A-2 x ATS4_HUMAN   ..

 Align seg 1/1  to: ATS4_HUMAN  from: 1  to: 837

      73 CCCCAACGCGGCCCCGCCGGCTCGGTGCCTGCCAGCGCCCCGGCCCGCAC 122
         |||   ||||||   ||||||            :::|||
       8 ProGlyArgGlyLeuAlaGlyArgTrpLeuTrpGlyAla..... 20

     123 GCCGCCCTGCCGCCTGCTT.........................C 142
             ||||||   ||||||                        |
      21 .GlnProCysLeuLeuLeuProIleValProLeuSerTrpLeuValTrpL 37

     143 TCGTCCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCGCGCC 192
         ||:::|||||||||||||      |||      |||:::|||
      37 euLeuLeuLeuLeuLeuAlaSerLeuLeuProSerAlaArg..... 50

     193 TGGGGGGCTGCTGCGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAAAA 242
                 |||:::|||
      51 ......LeuAlaSerProLeu..... 55

     243 AAATTTGGGAGTCCTGGCAGATGAAGACAATACATTGCAACAGAATAGCA 292

      55 ..... 55

     293 GCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAATCACACTGCCT 342
                                  ::::::||||||::::::|||
      56 .....ProArgGluGluGluIleValPhePro 64

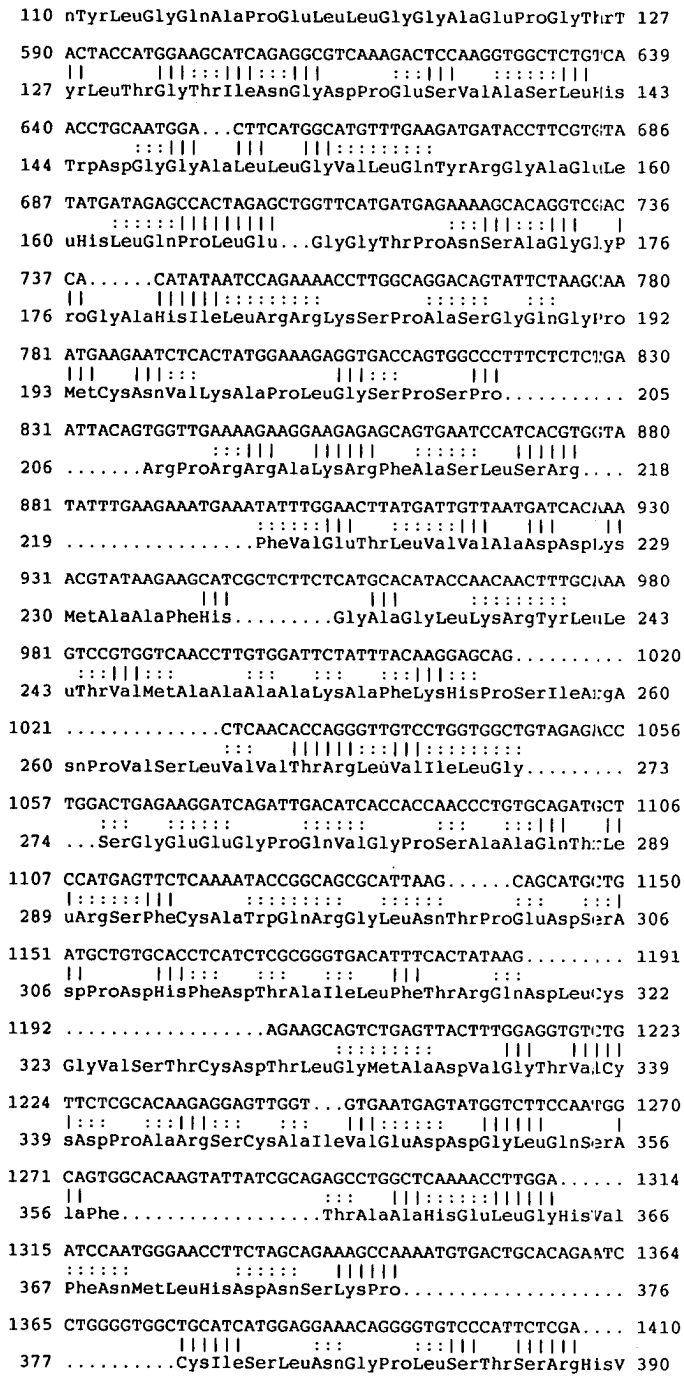
     343 TCAAGACTCATATATTACATCAACCAAGACTCGGAAAGCCCTTATCACGT 392
         ::::::|||         :::         |||   :::|||   :::::
      65 GluLysLeuAsnGlySerValLeuProGlySerGlyAlaProAlaArgLe 81

     393 TCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCATC 442
         :|||
      81 uLeu..... 82

     443 TGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTT 492
                         :::::::::||||||||||:::   ::::::|||
      83 .............CysArgLeuGlnAlaPheGlyGluThrLeuLeuLeu 94

     493 GACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCA 542
         :::|||      :::::::|||   ::::::   :::   :::   :::::
      95 GluLeuGluGlnAspSerGly...ValGlnValGluGlyLeuThrValGl 110

     543 CTACGAAAATGGGAAACCACAGTACTCTAAGGGTGGAGAG...CACTGTT 589
         :|||   :::      |||:::         |||:::|||         |
```


110 nTyrLeuGlyGlnAlaProGluLeuLeuGlyGlyAlaGluProGlyThrT 127
590 ACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCA 639
127 yrLeuThrGlyThrIleAsnGlyAspProGluSerValAlaSerLeuHis 143
640 ACCTGCAATGGA...CTTCATGGCATGTTTGAAGATGATACCTTCGTGTA 686
144 TrpAspGlyGlyAlaLeuLeuGlyValLeuGlnTyrArgGlyAlaGluLe 160
687 TATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACAGGTCGAC 736
160 uHisLeuGlnProLeuGlu...GlyGlyThrProAsnSerAlaGlyGlyP 176
737 CA.....CATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAA 780
176 roGlyAlaHisIleLeuArgArgLysSerProAlaSerGlyGlnGlyPro 192
781 ATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGA 830
193 MetCysAsnValLysAlaProLeuGlySerProSerPro..... 205
831 ATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGTA 880
206 ArgProArgArgAlaLysArgPheAlaSerLeuSerArg.... 218
881 TATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAA 930
219 PheValGluThrLeuValValAlaAspAspLys 229
931 ACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAA 980
230 MetAlaAlaPheHis.........GlyAlaGlyLeuLysArgTyrLeuLe 243
981 GTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAG..... 1020
243 uThrValMetAlaAlaAlaAlaLysAlaPheLysHisProSerIleArgA 260
1021 CTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACC 1056
260 snProValSerLeuValValThrArgLeuValIleLeuGly..... 273
1057 TGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAGATGCT 1106
274 ...SerGlyGluGluGlyProGlnValGlyProSerAlaAlaGlnThrLe 289
1107 CCATGAGTTCTCAAAATACCGGCAGCGCATTAAG......CAGCATGCTG 1150
289 uArgSerPheCysAlaTrpGlnArgGlyLeuAsnThrProGluAspSerA 306
1151 ATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAG..... 1191
306 spProAspHisPheAspThrAlaIleLeuPheThrArgGlnAspLeuCys 322
1192 AGAAGCAGTCTGAGTTACTTTGGAGGTGTCTG 1223
323 GlyValSerThrCysAspThrLeuGlyMetAlaAspValGlyThrValCy 339
1224 TTCTCGCACAAGAGGAGTTGGT...GTGAATGAGTATGGTCTTCCAATGG 1270
339 sAspProAlaArgSerCysAlaIleValGluAspAspGlyLeuGlnSerA 356
1271 CAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGA..... 1314
356 laPhe.................ThrAlaAlaHisGluLeuGlyHisVal 366
1315 ATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATC 1364
367 PheAsnMetLeuHisAspAsnSerLysPro..... 376
1365 CTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCATTCTCGA.... 1410
377 CysIleSerLeuAsnGlyProLeuSerThrSerArgHisV 390
1411 AAATTTTCAAAG 1422
390 alMetAlaProValMetAlaHisValAspProGluGluProTrpSerPro 406
1423 TGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTG 1472
407 CysSerAlaArgPheIleThrAspPheLeuAspAsnGlyTyrGlyHisCy 423
1473 CCTTTTCAACAGGCCA...ACAAAGCTATTTGAGCCC...ACGGAATGTG 1516
423 sLeuLeuAspLysProGluAlaProLeuHisLeuProValThrPheProG 440
1517 GAAATGGATACGTGGAAGCTGGGGAGGAGTGTGAT..... 1551
440 lyLysAspTyr...AspAlaAspArgGlnCysGlnLeuThrPheGlyPro 455
1552 TG 1553
456 AspSerArgHisCysProGlnLeuProProProCysAlaAlaLeuTrpCy 472
1554 TGGTTTTCATGTGGAATGCTATGGATTATGCTGTAAGAAATGTTCC...C 1600
472 sSerGlyHisLeuAsnGlyHisAlaMetCysGlnThrLysHisSerProT 489
1601 TCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCA 1650
489 rpAlaAspGlyThrProCysGlyProAlaGlnAlaCysMetGlyGlyArg 505
1651 TGTCTTTTT.....CAGCCACGAGG 1670
506 CysLeuHisMetAspGlnLeuGlnAspPheAsnIleProGlnAlaGlyGl 522
1671 GTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTA 1720
522 yTrpGlyProTrpGlyProTrpGlyAspCysSerArgThr.....CysG 537
1721 CTGGA.....GACTCTGGTCAGTGCCCACCAAATCTTCATAAG 1758
537 lyGlyGlyValGlnPheSerSerArgAspCysThrArgProValProArg 553
1759 CAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGA 1808
554 AsnGlyGlyLysTyrCys.....GluGlyArg..... 562
1809 GTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTG 1858
563 ArgThrArgPheArgSerCys..... 569
1859 CAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAG 1908
570 AsnThrGluAspCysPro 575
1909 AAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACA 1958
576 ThrGlySerAlaLeuThrPheArgGluGlu.....GlnCysAlaAlaTy 590
1959 T.....GATGTGTTCTGTGGATTC..... 1977
590 rAsnHisArgThrAspLeuPheLysSerPheProGlyProMetAspTrpV 607
1978 TTACTCTGT 1986
607 alProArgTyrThrGlyValAlaProGlnAspGlnCysLysLeuThrCys 623
1987 ACCAATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCAT 2036
623 623
2037 TCCAACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCC 2086
624 GlnAlaArgAlaLeu.....GlyTyrT 631

```
 2087 ATGTAGTTTTAGATGATGATACGGATGTGGGCTATGTAGAAGATGGAACG 2136
      ::   ||||||:::                    |||   |||||||||
  631 yrTyrValLeuGluPro..............ArgValValAspGlyThr 642

 2137 CCATGTGGCCCG......TCTATGATGTGTTTAGATCGGAAGTGCCTACA 2180
      ||||||:::|||      |||   :::|||::::::   :::|||:::::
  643 ProCysSerProAspSerSerSerValCysValGlnGlyArgCysIleHi 659

 2181 AATTCAAGCCCTAAATATGAGCAGCTGTCCA......CTCGATTCCAAGG 2224
      :                 ::::::|||         :::   ||||||
  659 s..................AlaGlyCysAspArgIleIleGlySerLysL 670

 2225 GTAAA..............GTCTGTTCGGGCCATGGGGTG...TGTAGT 2256
        |||              ||||||:::|||   |||      ||||||
  670 ysLysPheAspLysCysMetValCysGlyGlyAspGlySerGlyCysSer 686

 2257 AATGAAGCC 2265
      :::::::::
  687 LysGlnSer 689

seq_name: SwissProt_39:HR2A_TRIFL

[redacted]_documentation_block:
     HR2A_TRIFL     STANDARD;      PRT;   202 AA.
     P14530;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   HEMORRHAGIC METALLOPROTEINASE HR2A (EC 3.4.24.53) (TRIMERELYSIN II).
OS   Trimeresurus flavoviridis (Habu).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC   Viperidae; Crotalinae; Trimeresurus.
OX   NCBI_TaxID=88087;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Venom;
RX   MEDLINE=89327213; PubMed=2753880;
RA   Miyata T., Takeya H., Ozeki Y., Arakawa M., Tokunaga F., Iwanaga S.,
RA   Omori-Satoh T.;
RT   "Primary structure of hemorrhagic protein, HR2a, isolated from the
RT   venom of Trimeresurus flavoviridis.";
RL   J. Biochem. 105:847-853(1989).
CC   -!- FUNCTION: THIS PROTEIN IS A ZINC PROTEASE FROM SNAKE VENOM THAT
CC       ACTS IN HEMORRHAGE.
CC   -!- CATALYTIC ACTIVITY: CLEAVAGE OF 3-ASN-|-GLN-4, 10-HIS-|-LEU-11 AND
CC       14-ALA-|-LEU-15 IN THE INSULIN B CHAIN, AND THE BOND Z-GLY-PRO-|-
CC       LEU-GLY-PRO IN A SMALL MOLECULE SUBSTRATE OF MICROBIAL
CC       COLLAGENASE.
CC   -!- COFACTOR: BINDS ONE ZINC ION.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
     PIR; JX0074; HYTVH2.
     HSSP; P34179; 2AIG.
     MEROPS; M12.156; -.
     InterPro; IPR000130; -.
DR   InterPro; IPR001590; -.
DR   Pfam; PF01421; Reprolysin; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
KW   Hydrolase; Metalloprotease; Zinc; Venom.
FT   MOD_RES       1      1       PYRROLIDONE CARBOXYLIC ACID.
FT   METAL       143    143       ZINC (CATALYTIC) (PROBABLE).
FT   ACT_SITE    144    144       BY SIMILARITY.
FT   METAL       147    147       ZINC (CATALYTIC) (PROBABLE).
FT   METAL       153    153       ZINC (CATALYTIC) (PROBABLE).
FT   DISULFID    118    197
FT   DISULFID    159    181
FT   DISULFID    161    164
FT   VARIANT       1      1       MISSING (IN 50% OF THE CHAINS).
SQ   SEQUENCE   202 AA;  23038 MW;  8850C057CA21CEDB CRC64;
```

```
alignment_scores:
           Quality:  208.00            Length:    205
             Ratio:  1.651               Gaps:      7
 Percent Similarity:  61.463   Percent Identity: 27.805

alignment_block:
 US-09-634-252A-2 x HR2A_TRIFL   ..

 Align seg 1/1  to: HR2A_TRIFL  from: 1  to: 202

  895 AAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCA 944
      :::|||:::||||||   ||||||   ||||||      |||   |||::
    7 ArgTyrIleGluLeuAlaIleValValAspHisGlyMetTyrThrLysTy 23

  945 TCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACC 994
      :   |||:::          ::::::   :::   ::::::||||||
   23 rSerSerAsnPheLysLysIleArgLysArgValHisGlnMetValAsnA 40

  995 TTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTG 1044
        ::::::::::::|||:::      ||||||      :::::|||
   40 snIleAsnGluMetTyrArg...ProLeuAsnIleAlaIleThrLeuSer 55

 1045 GCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCC 1094
        :::::::::|||:::|||||||||   |||   :::   :::
   56 LeuLeuAspValTrpSerGluLysAspLeuIleThrMetGlnAlaValAl 72

 1095 TGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATT..... 1137
                :::   |||:::   :::|||:::   :::
   72 aProThrThrAlaArgLeuPheGlyAspTrpArgGluThrValLeuLeuL 89

 1138 .....AAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTT 1182
           |||:::|||   |||   ::::::|||::::::   ::::::|||
   89 ysGlnLysAspHis...AspHisAlaGlnLeuLeuThrAspIleAsnPhe 104

 1183 CACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCAC 1232
            :::      :::   :::|||:::||||||:::|||:::
  105 ThrGlyAsnThrIleGlyTrpAlaTyrMetGlyGlyMetCysAsnAlaLy 121

 1233 AAGAGGAGTTGGTGTG......AATGAGTATGGTCTTCCAATGGCAGTGG 1276
       ::::::||||||:::      ::::::   :::      :::   ||||
  121 sAsnSerValGlyIleValLysAspHisSerSerAsnValPheMetValA 138

 1277 CACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAA 1326
      ||   :::::::::::::::::::::::::::|||||||||::::::   :::
  138 laValThrMetThrHisGluIleGlyHisAsnLeuGlyMetGluHisAsp 154

 1327 CCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTG 1376
                 |||   ||||||   |||                 :::||
  155 Asp.........LysAspLysCysLysCysGlu............AlaCy 164

 1377 CATCATGGAGGAAACA...GGGGTGTCCCATTCTCGAAAATTTTCAAAGT 1423
      |||||||:::   :::   :::   :::   |||:::   ||||||   |
  164 sIleMetSerAlaValIleSerAspLysProSerLysLeuPheSerAspC 181

 1424 GCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGC 1473
      |||||         |||:::   ||||||   ::::::          |||
  181 ysSerLysAspTyrTyrGlnThrPheLeuThrAsnSerLysProGlnCys 197

 1474 CTTTTCAACAGGCCA 1488
      :::::::|||   |||
  198 IleIleAsnAlaPro 202
```

OM of: US-09-634-252A-2 to: PIR_68:* out_format : pfs

Date: Aug 15, 2001 9:58 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09634252/runat_14082001_142628_9823/app_query.fasta_1.2598
-DB=PIR_68 -QFMT=fastan -SUFFIX=n2p.rpr -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-[illegible]NS=human40.cdi -LIST=90 -DOCALIGN=200 -THR_SCORE=pct
-[illegible]_MAX=100 -THR_MIN=0 -ALIGN=30 -MODE=LOCAL -OUTFMT=pfs
-[illegible]RM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09634252_@CGN1_1_98 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
-WAIT -THREADS=1

Search information block:
Query: US-09-634-252A-2
Query length: 2499
Database: PIR_68:*
Database sequences: 219241
Database length: 76174552
Search time (sec): 109.950000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
pir2:I65967	+	1872.50	2272.20	2.1e-119	670	disintegrin-like metalloprotein
pir2:S38539	+	1297.00	1572.55	2.5e-80	524	disintegrin-like metalloprotein
pir2:S60257	+	966.00	1164.37	8.0e-58	903	meltrin alpha - mouse
pir2:G02390	+	826.00	994.54	2.6e-48	814	disintegrin-like metalloprotein
pir2:S28258	+	797.50	960.19	2.2e-46	776	androgen-regulated epididymal p
pir2:S55060	+	779.00	937.12	4.0e-45	825	fertilin alpha-II - crab-eating
pir2:S48160	+	778.50	938.96	4.2e-45	617	metalloproteinase (EC 3.4.24.-)
pir2:S28259	+	775.00	932.62	7.4e-45	789	androgen-regulated epididymal p
pir2:S55059	+	775.00	931.46	7.5e-45	905	fertilin alpha-I - crab-eating
pir2:S71949	+	774.00	932.90	8.6e-45	660	metalloproteinase 12 (EC 3.4.24
pir2:G02937	+	754.50	908.22	1.8e-43	735	fertilin beta - crab-eating mac
pir2:S24789	+	751.50	906.69	2.9e-43	571	jararhagin C precursor - jarara
pir2:T26644	+	746.50	895.52	6.6e-43	1042	hypothetical protein Y37D8A.13
[illegible]:JC4861	+	739.00	889.33	2.1e-42	734	fertillin beta chain - human
[illegible]:A55796	+	735.00	885.93	3.8e-42	616	ecarin precursor - saw-scaled v
[illegible]:JC7530	+	725.00	873.82	1.8e-41	610	vascular apoptosis-inducing pro
[illegible]:S55270	+	723.50	872.01	2.3e-41	609	catrocollastatin precursor - we
pir2:I48100	+	687.00	825.44	7.0e-39	777	ADAM 5 protein precursor - guin
pir2:A60385	+	685.50	823.10	8.9e-39	826	monocyte surface antigen MS2 pr
pir2:S47645	+	685.00	823.46	9.6e-39	736	tMEC I protein - crab-eating ma
pir2:S47656	+	681.50	818.97	1.7e-38	756	tMEC II protein - crab-eating m
pir2:T18900	+	667.00	799.34	1.6e-37	952	disintegrin and metalloproteina
pir2:S18968	+	658.50	790.21	6.1e-37	823	cyritestin precursor - mouse
pir2:I49281	+	651.50	784.34	1.8e-36	600	fertilin alpha precursor - mous
pir2:S48169	+	647.00	779.60	3.6e-36	549	metalloproteinase (EC 3.4.24.-)
pir2:A37877	+	610.00	736.83	1.1e-33	416	hemorrhagic protein HR1B - habu
pir2:A42972	+	595.00	718.28	1.2e-32	429	coagulation factor X activating
pir2:S41607	+	577.50	697.14	1.8e-31	419	atrolysin A (EC 3.4.24.1) - wes
pir2:I52361	+	572.50	686.34	4.2e-31	732	testicular metalloproteinase-li
pir2:I65253	+	566.50	679.24	1.1e-30	713	disintegrin-like testicular met
pir2:I48101	+	529.00	633.26	3.8e-28	735	ADAM 6 protein precursor - guin
pir2:I49283	+	503.00	605.28	2.1e-26	473	ADAM 4 protein precursor - mous
pir2:A43296	+	498.50	599.70	4.3e-26	478	atrolysin E (EC 3.4.24.44) - we
pir1:A30065	+	494.50	594.79	8.0e-26	480	trigramin precursor - Indian gr
pir2:S43125	+	482.00	579.53	5.7e-25	481	trimucin precursor - Chinese ha
pir2:JQ1301	+	475.50	571.66	1.6e-24	478	hemorrhagic proteinase (EC 3.4.
pir2:JC4342	+	474.50	570.39	1.8e-24	481	fibrinolytic proteinase (EC 3.4
pir2:JC4880	+	473.50	569.22	2.1e-24	478	fibrinolytic metalloproteinase
pir2:S23403	+	461.00	556.44	1.5e-23	357	sperm surface protein PH-30 bet
pir2:JX0266	+	412.50	501.82	2.8e-20	209	platelet aggregation inhibitor
pir1:HYSNFA	+	398.50	479.05	2.6e-19	411	fibrolase (EC 3.4.24.-) prec
pir2:JX0265	+	376.00	457.04	8.4e-18	216	platelet aggregation inhibit
pir2:S66260	+	373.50	448.65	1.3e-17	407	metalloproteinase (EC 3.4.24
pir1:HYRSAC	+	368.50	442.41	2.9e-17	414	atrolysin C (EC 3.4.24.42) p
pir2:S41609	+	368.50	442.41	2.9e-17	414	atrolysin C (EC 3.4.24.42) p
pir2:S41608	+	367.50	441.19	3.3e-17	414	atrolysin B (EC 3.4.24.41) p
pir2:S60258	+	360.50	441.22	9.2e-17	150	meltrin beta - mouse (fragme
pir2:S66129	+	338.50	400.83	3.3e-15	748	disintegrin (EC 3.4.24.-) pr
pir2:S60259	+	310.00	379.59	2.5e-13	151	meltrin gamma - mouse (fragm
pir2:S52477	+	303.00	360.24	8.2e-13	544	disintegrin (EC 3.4.24.-) pr
pir2:T00017	+	299.50	351.25	1.5e-12	951	gene ADAMTS-1 protein - mous
pir2:T13809	+	297.00	345.97	2.2e-12	1239	probable disintegrin (EC 3.
pir2:T37819	+	292.00	347.33	4.6e-12	512	probable zinc metallopeptida
pir2:S52920	+	273.50	325.13	8.2e-11	491	disintegrin (EC 3.4.24.-) -
pir2:S23402	+	269.00	324.12	1.6e-10	289	sperm surface protein PH-30
pir2:T37256	+	253.50	295.43	2.0e-09	922	metalloproteinase sup-17 - C
pir2:T25987	+	247.00	290.00	5.3e-09	686	hypothetical protein ZK154.7
pir2:T21371	+	246.00	279.08	6.8e-09	2165	hypothetical protein F25H8.
pir2:T18856	+	239.00	273.96	2.0e-08	1444	angiogenesis inhibitor homo
pir1:HYTV2	+	223.00	271.09	2.1e-07	201	trimerelysin II (EC 3.4.24.5
pir1:JQ1322	+	221.00	249.18	3.4e-07	2019	tenascin precursor - mouse
pir2:S46443	+	213.50	259.43	9.1e-07	203	adamalysin (EC 3.4.24.46) II
pir2:A61625	+	213.50	248.04	1.0e-06	782	tenascin-like protein precur
pir1:HYRSR	+	212.50	258.25	1.1e-06	202	ruberlysin (EC 3.4.24.48) -
pir2:T18517	+	210.50	240.74	1.7e-06	1205	procollagen N-endopeptidase
pir1:A32230	+	210.50	237.30	1.7e-06	1810	tenascin precursor - chicke
pir2:JC2550	+	209.50	254.59	1.7e-06	202	Ac1 metalloproteinase (EC 3.
pir1:S19694	+	209.00	235.78	2.2e-06	1746	tenascin precursor - pig
pir2:T00355	+	208.50	241.37	2.2e-06	837	hypothetical protein KIAA068
pir1:HYTVH2	+	208.00	252.76	2.1e-06	202	hemorrhagic protein HR2a (EC
pir1:A40701	+	208.00	228.53	2.7e-06	3566	tenascin-X precursor - huma
pir2:T16892	+	207.00	239.32	2.8e-06	860	hypothetical protein T19D2.1
pir2:S15111	+	206.50	251.02	2.7e-06	200	hemorrhagic factor LHFII (EC
pir2:A43902	+	204.50	238.67	4.1e-06	647	tenascin - eastern newt (fra
pir2:A32160	+	204.50	228.34	4.5e-06	2201	tenascin-C - human
pir3:JC7570	+	193.00	224.17	2.5e-05	685	Delta-4 protein - human
pir2:T09070	+	192.50	208.65	3.1e-05	4006	probable tenascin X - mouse
pir2:T31070	+	190.00	209.48	4.4e-05	2531	notch homolog - sea urchin
pir3:JC7569	+	188.50	218.67	5.0e-05	686	Delta-4 protein - mouse
pir2:T42629	+	188.50	203.51	5.8e-05	4135	tenascin-X - bovine
pir2:A32901	+	188.00	212.69	5.7e-05	1295	glp1 protein precursor - Ca
pir2:I48324	+	185.50	214.58	8.0e-05	722	DELTA-like 1 - mouse
pir2:S06434	+	185.00	208.21	9.2e-05	1429	homeotic protein lin-12 pre
pir2:T13954	+	185.00	207.39	9.3e-05	1574	MEGF6 protein - rat
pir2:T30201	+	185.00	204.00	9.6e-05	2352	Notch homolog protein - sea
pir2:T42218	+	184.50	207.01	0.00010	1531	slit-1 protein homolog - ra
pir2:T09059	+	183.00	203.08	0.0001	1964	notch4 - mouse
pir2:A46019	+	183.00	200.94	0.0001	2531	Notch-1 protein - mouse
pir2:T14271	+	183.00	200.02	0.0001	2825	Doc4 protein, stress-induce
pir2:I51280	+	182.50	221.79	0.0001	199	atroxase (EC 3.4.24.43) - we

seq_name: pir2:I65967

seq_documentation_block:
disintegrin-like metalloproteinase (EC 3.4.24.-), splice form 2 - human
N;Alternate names: ADAM11; MDC
C;Species: Homo sapiens (man)
C;Date: 29-May-1998 #sequence_revision 17-Mar-2000 #text_change 26-May-2000
C;Accession: I65967; S38539
R;Katagiri, T.; Harada, Y.; Emi, M.; Nakamura, Y.
Cytogenet. Cell Genet. 68, 39-44, 1995
A;Title: Human metalloprotease/disintegrin-like (MDC) gene: exon-intron organization
A;Reference number: I52965; MUID:95044425
A;Accession: I65967
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 7-670 <KAT>
A;Cross-references: GB:D31872; NID:g505090; PIDN:BAA06670.1; PID:d1007242; PID:g83668
R;Emi, M.; Katagiri, T.; Harada, Y.; Saito, H.; Inazawa, J.; Ito, I.; Kasumi, F.; Nak
Nature Genet. 5, 151-157, 1993
A;Title: A novel metalloprotease/disintegrin-like gene at 17q21.3 is somatically rear
A;Reference number: S38539; MUID:94073190
A;Accession: S38539

```
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-225,'N',227-495,'PQQGRAVWLPPLCQHLWSSSARGPGGRHQ' <EMI>
A;Cross-references: GB:D17390; NID:g452188; PIDN:BAA04213.1; PID:d1004732; PID:g484255
C;Comment: For an alternative splice form, see PIR:S38539.
C;Genetics:
A;Gene: GDB:MDC; ADAM11
A;Cross-references: GDB:230267; OMIM:155120
A;Map position: 17q21.3-17q21.3
A;Introns: 28/3; 57/2; 82/3; 105/1; 127/3; 152/3; 176/3; 232/2; 260/3; 291/1; 308/2; 341
C;Superfamily: disintegrin homology
C;Keywords: alternative splicing; hydrolase; metalloproteinase
F;344-427/Domain: disintegrin homology <DIS>


alignment_scores:
           Quality: 1872.50              Length:    689
             Ratio:  3.650                Gaps:       13
 Percent Similarity:  74.456   Percent Identity:  51.959

[redacted]ment_block:
[redacted]9-634-252A-2 x I65967   ..

Align seg 1/1  to: I65967  from: 1  to: 670

    502 CTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTAC...GA 548
        |||::::::   ||||||||||||:::|||||||||   |||:::   ::
      4 LeuSerHisGlnLeuLeuSerSerGlnTyrValGluArgHisPheSerAr 20

    549 AAATGGGAAACCACAGTACTCTAAGGGT...GGAGAGCACTGTTACTACC 595
        ::::|||      |||:::|||  |||   |||:::||||||||||||:
     20 gGluGlyThrThrGlnHisSerThrGlyAlaGlyAspHisCysTyrTyrG 37

    596 ATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGC 645
        ::|||:::::|||||         |||   :::|||||||||||||||||
     37 lnGlyLysLeuArgGlyAsnProHisSerPheAlaAlaLeuSerThrCys 53

    646 AATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGA 695
        :::|||||||||||||:::|||:::|||   :::::::::|||:::::||
     54 GlnGlyLeuHisGlyValPheSerAspGlyAsnLeuThrTyrIleValGl 70

    696 GCCACTAGAGCTGGTT......CATGATGAGAAAAGCACAGGTCGACCAC 739
        ||||   |||::::::                :::          ||||
     70 uProGlnGluValAlaGlyProTrpGlyAlaProGlnGlyProLeuProH 87

    740 ATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAAT 789
        ||:::|||   :::|||
     87 isLeuIleTyrArgThr..................................  92

    790 CTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTG 839
                                |||:::|||   :::
     93 ..........................ProLeuLeuProAspProLeuGl 100

     40 GTTGAAA.....................................------ 846
           :::
    100 yCysArgGluProGlyCysLeuPheAlaValProAlaGlnSerAlaProP 117

    847 ..................AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT 879
                         ||||||||||||   |||:::      ||||||
    117 roAsnArgProArgLeuArgArgLysArgGlnValArg......ArgGly 131

    880 .........ATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAA 920
                :::   :::|||   ||||||:::||||||:::::::::||
    132 HisProThrValHisSerGluThrLysTyrValGluLeuIleValIleAs 148

    921 TGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAACA 970
        |||||||:::   :::::::::   |||:::|||   :::   |||:::|
    148 nAspHisGlnLeuPheGluGlnMetArgGlnSerValValLeuThrSerA 165

    971 ACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAG 1020
        ||||||||||||||||||||||||||:::|||   ||||||||||||||
    165 snPheAlaLysSerValValAsnLeuAlaAspValIleTyrLysGluGln 181

   1021 CTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGA 1070
        ||||||||||||:::|||||||||||||:::|||||||||::::::   ||
    182 LeuAsnThrArgIleValLeuValAlaMetGluThrTrpAlaAspGlyAs 198

   1071 TCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAA 1120
        |:::|||::::::      :::   ::::::   |||   ::::::
    198 pLysIleGlnValGlnAspAspLeuLeuGluThrLeuAlaArgLeuMetV 215

   1121 AATACCGGCAGCGC...ATTAAGCAGCATGCTGATGCTGTGCACCTCATC 1167
          ||||||::::::   :::   :::   :::||||||:::|||||||::
    215 alTyrArgArgGluGlyLeuProGluProSerAspAlaThrHisLeuPhe 231

   1168 TCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGG 1217
        |||      ||||||:::        |||:::  :::|||   |||||
    232 SerGlyArgThrPheGlnSerThrSerSerGlyAlaAlaTyrValGlyGl 248

   1218 TGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAA 1267
        |:::||||||   :::::::|||   ||||||||||||||||||
    248 yIleCysSerLeuSerHisGlyGlyGlyValAsnGluTyrGlyAsnMetG 265

   1268 TGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATC 1317
          |||:::|||   :::|||:::|||:::|||:::||||||||||||:::
    265 lyAlaMetAlaValThrLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMet 281

   1318 CAATGG...GAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATC 1364
        :::|||   :::      |||              |||   |||   :::
    282 MetTrpAsnLysHisArgSerSerAlaGlyAspCysLysCysProAspIl 298

   1365 CTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAAT 1414
          |||   |||||||||||||||:::||||||             ||||||
    298 eTrpLeuGlyCysIleMetGluAspThrGlyPheTyrLeuProArgLysP 315

   1415 TTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGT 1464
        |||||:::|||||||||   ||||||::::::|||||||||:::||||||
    315 heSerArgCysSerIleAspGluTyrAsnGlnPheLeuGlnGluGlyGly 331

   1465 GGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATG 1514
        |||:::|||||||||||:::|||   ||||||::::::|||   |||||
    332 GlySerCysLeuPheAsnLysProLeuLysLeuLeuAspProProGluCy 348

   1515 TGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATG 1564
        ||||||||||:::|||||||||||||||||||||||||||||||||||
    348 sGlyAsnGlyPheValGluAlaGlyGluGluCysAspCysGlySerValG 365

   1565 TGGAATGCTAT.........GGATTATGCTGTAAGAAATGTTCCCTCTCC 1605
          ||||||           |||   |||||||||||||||:::|||:::
    365 lnGluCysSerArgAlaGlyGlyAsnCysCysLysLysCysThrLeuThr 381

   1606 AACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCT 1655
        :::   |||   ||||||||||||   ||||||::::::      |||
    382 HisAspAlaMetCysSerAspGlyLeuCysCysArgArg......CysLy 396

   1656 TTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATA 1705
          ::::::||||||||   :::||||||:::||||||||||||||||||
    396 sTyrGluProArgGlyValSerCysArgGluAlaValAsnGluCysAspI 413

   1706 TTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCAT 1755
        ||:::|||   ||||||||||||||:::||||||||||||||||||||||
    413 leAlaGluThrCysThrGlyAspSerSerGlnCysProProAsnLeuHis 429

   1756 AAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGG 1805
        |||   |||||||||   |||:::::::::||||||||||||||||:::||
    430 LysLeuAspGlyTyrTyrCysAspHisGluGlnGlyArgCysTyrGlyGl 446

   1806 CGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGG 1855
        |:::|||||||||||||||:::|||||||||   :::||||||      |
    446 yArgCysLysThrArgAspArgGlnCysGlnValLeuTrpGlyHisAlaA 463

   1856 CTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACT 1905
        |||||      |||:::|||||||||||||||||||||:::|||||||||
    463 laAla......AspArgPheCysTyrGluLysLeuAsnValGluGlyThr 477

   1906 GAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAA 1955
```

```
     |||:::|||:::||||||:::   |||:::   |||:::||||||||||||
 478 GluArgGlySerCysGlyArgLysGlySerGlyTrpValGlnCysSerLy 494

1956 ACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTC 2005
     |:::||||||:::|||||||||||||||||||:::|||:::::::   ||||
 494 sGlnAspValLeuCysGlyPheLeuLeuCysValAsnIleSerGlyAlaP 511

2006 CACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCAT 2055
     |||||:::|||:::|||   |||:::|||      :::::::||||||||
 511 roArgLeuGlyAspLeuValGlyAspIleSerSerValThrPheTyrHis 527

2056 CAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGA 2105
     |||||||:::   :::||||||   |||:::||||||   |||   |||
 528 GlnGlyLysGluLeuAspCysArgGlyGlyHisValGlnLeuAlaAspGl 544

 106 TACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGA 2155
      :::|||::::::|||||||||||||||||||   |||||||||:::|||:
 544 ySerAspLeuSerTyrValGluAspGlyThrAlaCysGlyProAsnMetL 561

2156 TGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGC 2205
     ::|||||||||:::::||||||      :::|||:::|||:::|||:::
 561 euCysLeuAspHisArgCysLeuProAlaSerAlaPheAsnPheSerThr 577

2206 TGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAG 2255
     ||||||   :::::::::   :::::|||||   ||||||||||||||
 578 CysProGlySerGlyGluArgArgIleCysSerHisHisGlyValCysSe 594

2256 TAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCA 2305
     |||||||:::   |||||||||:::      |||:::|||   ||||'||
 594 rAsnGluGlyLysCysIleCysGlnProAspTrpThrGlyLysAspCysS 611

2306 GTATCCGGGATCCAGTTAGGAACCTTCACCCCCCC...AAGGATGAAGGA 2352
     |||||:::::::|||:::   :::      |||      :::   |||
 611 erIleHisAsnProLeuProThrSerProProThrGlyGluThrGluArg 627

2353 CCCAAGGGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGGTGC 2402
        ||||||||||||:::||||||:::||||||||||||||||||||| ||
 628 TyrLysGlyProSerGlyThrAsnIleIleIleGlySerIleAlaGlyAl 644

2403 CATCCTGGTAGCAGCTATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAA 2452
     |:::|||||||||||||||||||||||||||||||||||||||||||| ||
 644 aValLeuValAlaAlaIleValLeuGlyGlyThrGlyTrpGlyPheLysA 661

2453 ATGTCAAGAAGAGAAGG 2469
     ||:::::::::   |||
 661 snIleArgArgGlyArg 666

Name: pir2:S38539

Seq_documentation_block:
disintegrin-like metalloproteinase (EC 3.4.24.-), splice form 1 - human
N;Alternate names: ADAM11; MDC
C;Species: Homo sapiens (man)
C;Date: 07-Oct-1994 #sequence_revision 17-Nov-1995 #text_change 26-May-2000
C;Accession: S38539; I52965
R;Emi, M.; Katagiri, T.; Harada, Y.; Saito, H.; Inazawa, J.; Ito, I.; Kasumi, F.; Nakamu
Nature Genet. 5, 151-157, 1993
A;Title: A novel metalloprotease/disintegrin-like gene at 17q21.3 is somatically rearran
A;Reference number: S38539; MUID:94073190
A;Accession: S38539
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-524 <EMI>
A;Cross-references: GB:D17390; NID:g452188; PIDN:BAA04213.1; PID:d1004732; PID:g484255
R;Katagiri, T.; Harada, Y.; Emi, M.; Nakamura, Y.
Cytogenet. Cell Genet. 68, 39-44, 1995
A;Title: Human metalloprotease/disintegrin-like (MDC) gene: exon-intron organization and
A;Reference number: I52965; MUID:95044425
A;Accession: I52965
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 7-225,'D',227-524 <KAT>
A;Cross-references: GB:D31872; NID:g505090; PIDN:BAA06671.1; PID:d1007243; PID:g836684
C;Comment: For an alternative splice form, see PIR:I65967.
C;Genetics:
A;Gene: GDB:MDC; ADAM11
A;Cross-references: GDB:230267; OMIM:155120
A;Map position: 17q21.3-17q21.3
A;Introns: 28/3; 57/2; 82/3; 105/1; 127/3; 152/3; 176/3; 232/2; 260/3; 291/1; 308/2;
C;Superfamily: disintegrin homology
C;Keywords: alternative splicing; hydrolase; metalloproteinase
F;344-427/Domain: disintegrin homology <DIS>


alignment_scores:
           Quality: 1297.00              Length:    517
             Ratio: 3.515                  Gaps:     12
 Percent Similarity: 71.373   Percent Identity: 49.516

alignment_block:
 US-09-634-252A-2 x S38539   ..

 Align seg 1/1  to: S38539  from: 1  to: 524

    502 CTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTAC...GA 548
        |||::::::   ||||||||||||:::|||||||||   |||:::   ::
      4 LeuSerHisGlnLeuLeuSerSerGlnTyrValGluArgHisPheSerAr 20

    549 AAATGGGAAACCACAGTACTCTAAGGGT...GGAGAGCACTGTTACTACC 595
        ::::|||      |||:::|||   |||   |||:::|||||||||||:
     20 gGluGlyThrThrGlnHisSerThrGlyAlaGlyAspHisCysTyrTyrG 37

    596 ATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGC 645
        ::|||::::::||||||         |||   :::|||||||||||||||
     37 lnGlyLysLeuArgGlyAsnProHisSerPheAlaAlaLeuSerThrCys 53

    646 AATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGA 695
        :::|||||||||||:::|||:::|||   ::::::::::|||:::::||
     54 GlnGlyLeuHisGlyValPheSerAspGlyAsnLeuThrTyrIleValGl 70

    696 GCCACTAGAGCTGGTT......CATGATGAGAAAAGCACAGGTCGACCAC 739
        ||||   |||::::::                  :::         ||||
     70 uProGlnGluValAlaGlyProTrpGlyAlaProGlnGlyProLeuProH 87

    740 ATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAAT 789
        ||:::|||   :::|||
     87 isLeuIleTyrArgThr..............................  92

    790 CTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTG.839
                                     |||:::|||   :::
     93 ..........................ProLeuLeuProAspProLeuGl 100

    840 GTTGAAA.....................................  846
           :::
    100 yCysArgGluProGlyCysLeuPheAlaValProAlaGlnSerAlaProP 117

    847 .................AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT 879
                        ||||||||||||   |||:::      ||||||
    117 roAsnArgProArgLeuArgArgLysArgGlnValArg......ArgGly 131

    880 .........ATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAA 920
                ::::   :::|||   ||||||:::||||||:::::::::||
    132 HisProThrValHisSerGluThrLysTyrValGluLeuIleValIleAs 148

    921 TGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAACA 970
        |||||||:::   :::::::::   |||:::|||   :::   |||:::|
    148 nAspHisGlnLeuPheGluGlnMetArgGlnSerValValLeuThrSerA 165

    971 ACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAG 1020
        ||||||||||||||||||||||||||:::|||   ||||||||||||||
    165 snPheAlaLysSerValValAsnLeuAlaAspValIleTyrLysGluGln 181

   1021 CTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGA 1070
        ||||||||||||:::||||||||||||:::|||||||||::::::   ||
    182 LeuAsnThrArgIleValLeuValAlaMetGluThrTrpAlaAspGlyAs 198
```

```
1071 TCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAA 1120
 198 pLysIleGlnValGlnAspAspLeuLeuGluThrLeuAlaArgLeuMetV 215

1121 AATACCGGCAGCGC...ATTAAGCAGCATGCTGATGCTGTGCACCTCATC 1167
 215 alTyrArgArgGluGlyLeuProGluProSerAsnAlaThrHisLeuPhe 231

1168 TCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGG 1217
 232 SerGlyArgThrPheGlnSerThrSerSerGlyAlaAlaTyrValGlyGl 248

1218 TGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAA 1267
 248 yIleCysSerLeuSerHisGlyGlyGlyValAsnGluTyrGlyAsnMetG 265

1268 TGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATC 1317
 265 lyAlaMetAlaValThrLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMet 281

1318 CAATGG...GAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATC 1364
 282 MetTrpAsnLysHisArgSerSerAlaGlyAspCysLysCysProAspIl 298

1365 CTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAAT 1414
 298 eTrpLeuGlyCysIleMetGluAspThrGlyPheTyrLeuProArgLysP 315

1415 TTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGT 1464
 315 heSerArgCysSerIleAspGluTyrAsnGlnPheLeuGlnGluGlyGly 331

1465 GGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATG 1514
 332 GlySerCysLeuPheAsnLysProLeuLysLeuLeuAspProProGluCy 348

1515 TGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATG 1564
 348 sGlyAsnGlyPheValGluAlaGlyGluGluCysAspCysGlySerValG 365

1565 TGGAATGCTAT.....GGATTATGCTGTAAGAAATGTTCCCTCTCC 1605
 365 lnGluCysSerArgAlaGlyGlyAsnCysCysLysLysCysThrLeuThr 381

1606 AACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCT 1655
 382 HisAspAlaMetCysSerAspGlyLeuCysCysArgArg......CysLy 396

1656 TTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATA 1705
 396 sTyrGluProArgGlyValSerCysArgGluAlaValAsnGluCysAspI 413

1706 TTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCAT 1755
 413 leAlaGluThrCysThrGlyAspSerSerGlnCysProProAsnLeuHis 429

1756 AAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGG 1805
 430 LysLeuAspGlyTyrTyrCysAspHisGluGlnGlyArgCysTyrGlyGl 446

1806 CGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGG 1855
 446 yArgCysLysThrArgAspArgGlnCysGlnValLeuTrpGlyHisAlaA 463

1856 CTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACT 1905
 463 laAla......AspArgPheCysTyrGluLysLeuAsnValGluGlyThr 477

1906 GAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAA 1955
 478 GluArgGlySerCysGlyArgLysGlySerGlyTrpValGlnCysSerLy 494

1956 A 1956
```

```
     494 s 494

seq_name: pir2:S60257

seq_documentation_block:
meltrin alpha - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 20-Jun-2000
C;Accession: S60257
R;Yagami-Hiromasa, T.; Sato, T.; Kurisaki, T.; Kamijo, K.; Nabeshima, Y.I.; Fujisawa-
Nature 377, 652-656, 1995
A;Title: A metalloprotease-disintegrin participating in myoblast fusion.
A;Reference number: S60257; MUID:96026308
A;Accession: S60257
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-903 <YAG>
A;Cross-references: EMBL:D50411; NID:g1054586; PIDN:BAA08912.1; PID:g1054587
C;Superfamily: mouse meltrin alpha; disintegrin homology
F;421-503/Domain: disintegrin homology <DIS>
F;349/Active site: Glu #status predicted


alignment_scores:
            Quality:  966.00              Length:    822
              Ratio:  2.055                 Gaps:     32
 Percent Similarity:  57.178   Percent Identity: 31.630

alignment_block:
 US-09-634-252A-2 x S60257   ..

 Align seg 1/1  to: S60257  from: 1  to: 903

    103 GCCAGCGCCCCGGCC...CGCACGCCGCCCTGCCGCCTGCTTCTCGTCCT 149
      2 AlaGluArgProAlaArgArgAlaProProAlaArgAlaLeuLeuLeuAl 18

    150 TCTC.........CTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCGCG 190
     18 aLeuAlaGlyAlaLeuLeuAlaProArgAlaAlaArgGlyMet...SerL 34

    191 CCTGGGGGGCTGCTGCGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAA 240
     34 euTrp..................................................... 35

    241 AAAAATTTGGGAGTCCTGGCAGATGAAGACAATACATTGCAACAGAATAG 290
     36 .....................................AspGlnArgGl 39

    291 CAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAATCACACTGC 340
     39 yAlaTyrGluValAlaArgAlaSerLeuLeuSerLysAspProGlyIleP 56

    341 CTTCAAGACTCATATATTACATCAACCAAGACTCGGAAAGCCCTTATCAC 390
     56 roGlyGlnSerIle.............................................. 60

    391 GTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCA 440
     61 ............ProAlaLysAsp.....................Hi 65

    441 TCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTC 490
     65 sProAspValLeuThrValGlnLeuGlnLeuGluSerArgAspLeuIleL 82

    491 TTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATT 540
     82 euSerLeuGluArgAsnGluGlyLeuIleAlaAsnGlyPheThrGluThr 98

    541 CACTACGAAAATGGGAAACCACAGTACTCT.........AAGGGTGGAGA 581
     99 HisTyrLeuGlnAspGlyThrAspValSerLeuThrArgAsnHisThrAs 115
```


582 GCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGG 631
115 pHisCysTyrTyrHisGlyHisValGlnGlyAspAlaAlaSerValVaLS 132
632 CTCTGTCAACCTGCAATGGACTTCATGGC......ATGTTTGAAGATGAT 675
132 erLeuSerThrCysSerAspLeuArgGlyLeuIleMetPheGluAsnLys 148
676 ACCTTCGTGTATATGATAGAGCCACTA.....GA 704
149 Thr......TyrSerLeuGluProMetLysAsnThrThrAspSerTyrLy 163
705 GCTGGTTCATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAA 754
163 sLeuValProAlaGluSerMetThrAsn.....IleGlnGlyL 176
755 CCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGA 804
176 euCysGlySerGlnHisAsnLys......SerAsnLeuThrMetGlu... 189
805 GGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAAGCAA 854
190 ...AspValSerProGlyThrSerGlnMet..... 198
855 GAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATTTGG 904
199 .ArgAlaArgArgHisLysArgGluThrLeuLysMetThrLysTyrValG 215
905 AACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCT 954
215 luLeuValIleValAlaAspAsnArgGluPheGlnArgGlnGlyLysAsp 231
955 CATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTC 1004
232 LeuGluLysValLysGlnArgLeuIleGluIleAlaAsnHisValAspGly 248
1005 TATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGA 1054
248 sPheTyrArg...ProLeuAsnIleArgIleValLeuValGlyValGluV 264
1055 CCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAGATG 1104
264 alTrpAsnAspIleAspLysCysSerIleSerGlnAspProPheThrArg 280
1105 CTCCATGAGTTCTCAAAATACCGGCAGCGCATTAAG.....CA 1142
281 LeuHisGluPheLeuAspTrpArg...LysIleLysLeuLeuProArgLy 296
1143 GCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGA 1192
296 sSerHisAspAsnAlaGlnLeuIleSerGlyValTyrPheGlnGlyThrT 313
1193 GAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCT......CGCACAAGA 1236
313 hrIleGlyMetAlaProIleMetSerMetCysThrAlaGluGlnSerGly 329
1237 GGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACAAGTATT 1286
330 GlyValValMetAspHisSerAspSerProLeuGlyAlaAlaValThrLe 346
1287 ATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCA 1336
346 uAlaHisGluLeuGlyHisAsnPheGlyMetAsnHisAspThrLeuGluA 363
1337 GAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCATGGAG 1386
363 rgGlyCysSerCysArgMetAlaAlaGluLysGlyGlyCysIleMetAsn 379
1387 ...GAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTT 1433
380 ProSerThrGlyPheProPheProMetValPheSerSerCysSerArgLy 396
1434 GGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACA 1483
396 sAspLeuGluAlaSerLeuGluLysGlyMetGlyMetCysLeuPheAsnL 413
1484 GGCCA......ACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATAC 1527
413 euProGluValLysGlnAlaPheGlyGlyArgLysCysGlyAsnGlyTyr 429
1528 GTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTATGG 1577
430 ValGluGluGlyGluGluCysAspCysGlyGluProGluGluCysThrAs 446
1578 ATTATGCTGTAAG......AAATGTTCCCTCTCCAACGGGGCTCACTGCA 1621
446 nArgCysCysAsnAlaThrThrCysThrLeuLysProAspAlaValCysA 463
1622 GCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGG 1671
463 laHisGlyGlnCysCysGluAsp......CysGlnLeuLysProProGly 477
1672 TATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTAC 1721
478 ThrAlaCysArgGlySerSerAsnSerCysAspLeuProGluPheCysTh 494
1722 TGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATG 1771
494 rGlyThrAlaProHisCysProAlaAsnValTyrLeuHisAspGlyHisP 511
1772 CATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGA 1821
511 roCysGlnGlyValAspGlyTyrCysTyrAsnGlyIleCysGlnThrHis 527
1822 GACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAA 1871
528 GluGlnGlnCysValThrLeuTrpGlyProGlyAlaLysProAlaProGl 544
1872 GTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCG 1921
544 yIleCysPheGluArgValAsnSerAlaGlyAspProTyrGlyAsnCysG 561
1922 GGAAGGATGGA...GACCGGTGGATTCAGTGCAGCAAACATGATGTGTTC 1968
561 lyLysAspSerLysSerAlaPheAlaLysCysGluLeuArgAspAlaLys 577
1969 TGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCA 2018
578 CysGlyLysIleGlnCysGlnGlyGlyAlaSerArgProValIleGly.. 593
2019 ACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAA......GGCC 2062
594 .ThrAsnAlaValSerIleGluThrAsnIleProGlnGlnGluGlyGlyA 610
2063 GGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGAT...ACG 2109
610 rgIleLeu...CysArgGlyThrHisValTyrLeuGlyAspAspMetPro 625
2110 GATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTG 2159
626 AspProGlyLeuValLeuAlaGlyThrLysCysAlaGluGlyLysIleCy 642
2160 TTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTC 2209
642 sLeuAsnArgArgCysGlnAsnIleSerValPheGlyValHisLysCysA 659
2210 CACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAAT 2259
659 laMetGln.....CysHisGlyArgGlyValCysAsnAsn 670
2260 GAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTAT 2309
671 ArgLysAsnCysHisCysGluAlaHisTrpAla..... 681
2310 CCGGGATCCAGTTAGGAACCTTCACCCCCCC..... 2340

```
                                 ||||||
    682 ........................ProProPheCysAspLysPheGlyP 690

   2341 ...........AAGGATGAAGGACCC...AAGGGTCCTAGTGCCACCAAT 2376
                   |||:::||||||   :::      :::      :::
    690 heGlyGlySerThrAspSerGlyProIleArgGlnAlaAspAsnGlnGly 706

   2377 CTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCTATTGTCCT 2426
        |||   :::|||               ||||||||::   ::   ||
    707 LeuThrValGly..............IleLeuValSerIleLeuCysLe 718

   2427 TGGGGGCACAGGCTGG 2442
        |   ::::::|||:::
    718 uLeuAlaAlaGlyPhe 723

seq_name: pir2:G02390

seq_documentation_block:
disintegrin-like metalloproteinase MDC15 (EC 3.4.24.-) - human
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 31-Dec-2000
C;Accession: G02390; PC4263
A;Authors: Wren, B.; Raines, E.W.; Ross, R.
A;Submitted to the EMBL Data Library, January 1996
A;Reference number: H01157
A;Accession: G02390
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-814 <HER>
A;Cross-references: EMBL:U46005; NID:g1335871; PIDN:AAC51112.1; PID:g1335872
R;McKie, N.; Edwards, T.; Dallas, D.J.; Houghton, A.; Stringer, B.; Graham, R.; Russell,
Biochem. Biophys. Res. Commun. 230, 335-339, 1997
A;Title: Expression of members of a novel membrane linked metalloproteinase family (ADAM
A;Reference number: PC4263; MUID:97168971
A;Accession: PC4263
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-461 <MCK>
A;Experimental source: articular chondrocyte
C;Comment: This protein is a membrane bound protein and involved in cell/cell and cell/m
C;Superfamily: mouse meltrin alpha; disintegrin homology
C;Keywords: hydrolase; metalloproteinase; zinc
F;420-503/Domain: disintegrin homology <DIS>
F;348,352,358/Binding site: zinc (His) #status predicted
F;349/Active site: Glu #status predicted


alignment_scores:
            Quality:  826.00            Length:    863
              Ratio:  1.796               Gaps:     32
 Percent Similarity:  53.302   Percent Identity: 26.767

[redacted]ment_block:
[redacted]9-634-252A-2 x G02390   ..

 Align seg 1/1  to: G02390  from: 1  to: 814

     91 GGCTCGGTGCCTGCCAGCGCCCCGGCCCGCACGCCGCCCTGCCGCCTGCT 140
        |||   :::   |||:::::|||      :::   |||
     10 GlyLeuLeuGlyAlaGlySerProLeuProSerTrpPro.......... 22

    141 TCTCGTCCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCGCG 190
                      ||||||   :::::::::::::::::   :::|
     23 ..............LeuProAsnIleGlyGlyThrGluGluGlnGlnA 34

    191 CCTGGGGGGCTGCTGCGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAA 240
        ||   :::      ||||||     |||
     34 laGluSerGluLysAlaProArgGluProLeu.................. 44

    241 AAAAATTTGGGAGTCCTGGCAGATGAAGACAATACATTGCAACAGAATAG 290
                           |||   ::::::||||||:::::
     45 ......................GluProGlnValLeuGlnAspAspLe 53

    291 CAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAATCACACTGC 340
                |||||   ::::::::::|||   ::::::      |
     53 uPro......IleSerLeuLysLysValLeuGlnThrSerLeuProGluP 68

    341 CTTCAAGACTCATATATTACATCAACCAAGACTCGGAAAGCCCTTATCAC 390
        ||   |||:::
     68 roLeuArgIle..................................... 71

    391 GTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCA 440

     71 ..................................................... 71

    441 TCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTC 490
                           ::::::|||      |||::::::   ||||
     72 ..................LysLeuGluLeuAspGlyAspSerHisIleL 82

    491 TTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATT 540
        ||:::|||:::   |||:::   |||:::   :::      :::   :::
     82 euGluLeuLeuGlnAsnArgGluLeuValProGlyArgProThrLeuVal 98

    541 CACTACGAA...AATGGGAAACCACAGTACTCTAAGGGTGGA......GA 581
           |||:::   :::|||            |||:::|||         ||
     99 TrpTyrGlnProAspGlyThrArgValValSerGluGlyHisThrLeuGl 115

    582 GCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGG 631
        |:::|||   |||:::|||   :::|||||||         |||   |||:
    115 uAsnCysCysTyrGlnGlyArgValArgGlyTyrAlaGlySerTrpValS 132

    632 CTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATGATACCTTC 681
        :::::   ||||||:::||||||:::|||
    132 erIleCysThrCysSerGlyLeuArgGly..................... 141

    682 GTGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACAGG 731
                           |||   ::::::      |||:::|||
    142 ..................LeuValValLeuThrProGluArgSer..... 150

    732 TCGACCACATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAA 781

    150 ..................................................... 150

    782 TGAAGAATCTCACTATGGAAAGAGGT............GACCAGTGGCCC 819
                   |||:::|||:::|||            :::      |||
    151 ........TyrThrLeuGluGlnGlyProGlyAspLeuGlnGlyProPro 164

    820 TTTCTCTCTGAATTACAG.......................... 837
        ::::::|||::::::|||
    165 IleIleSerArgIleGlnAspLeuHisLeuProGlyHisThrCysAlaLe 181

    838 ....TGGTTGAAAAGA.............................A 850
            |||   :::
    181 uSerTrpArgGluSerValHisThrGlnThrProProGluHisProLeuG 198

    851 GGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATAT 900
          :::|||   ::::::      |||   ::      |||   |||
    198 lyGlnArgHisIleArgArgArgArgAspValValThrGluThrLysThr 214

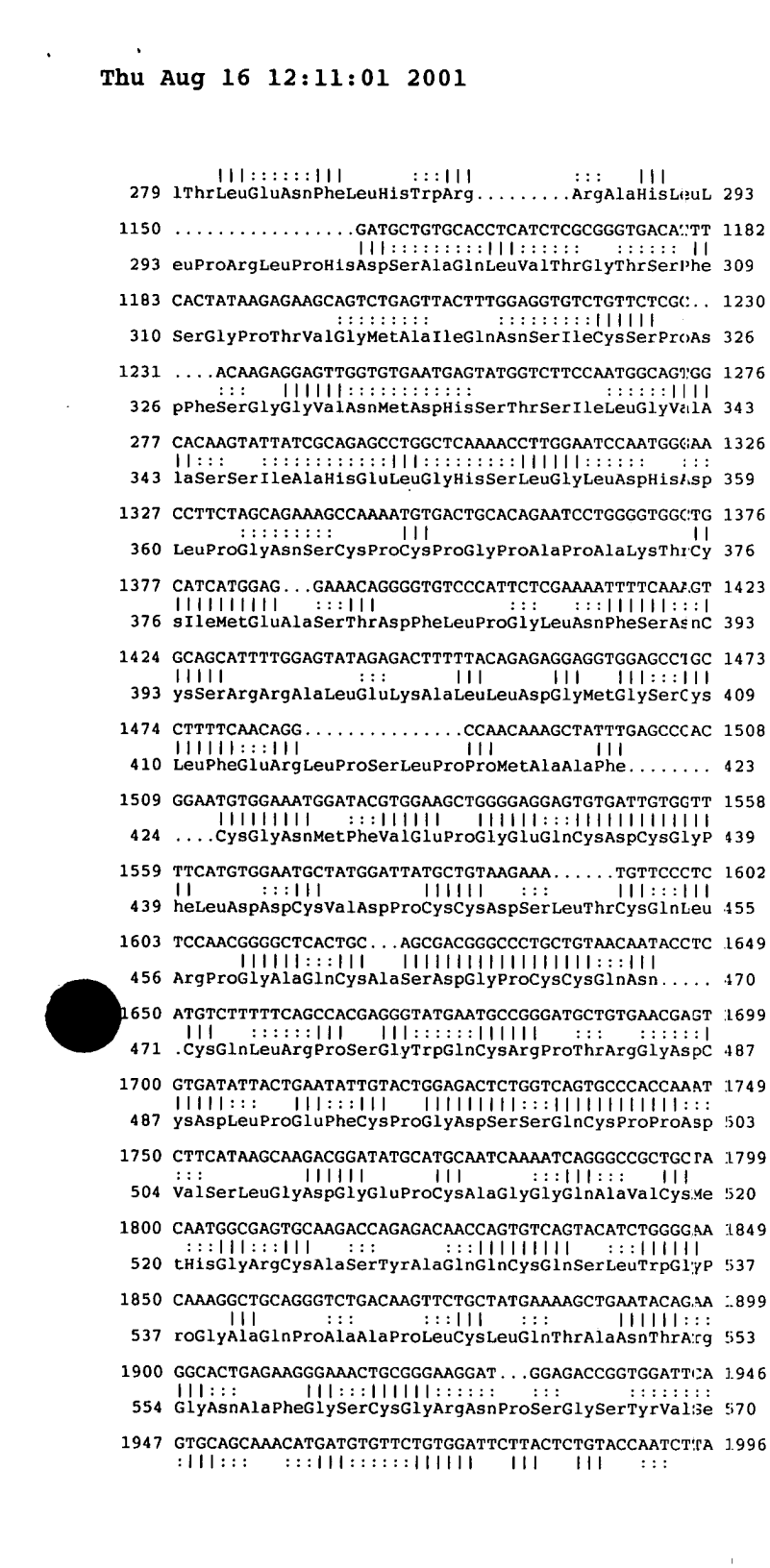
    901 TTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTC 950
        :::||||||:::||||||   ||||||:::      :::|||:::|||
    215 ValGluLeuValIleValAlaAspHisSerGluAlaGlnLysTyrArg.. 230

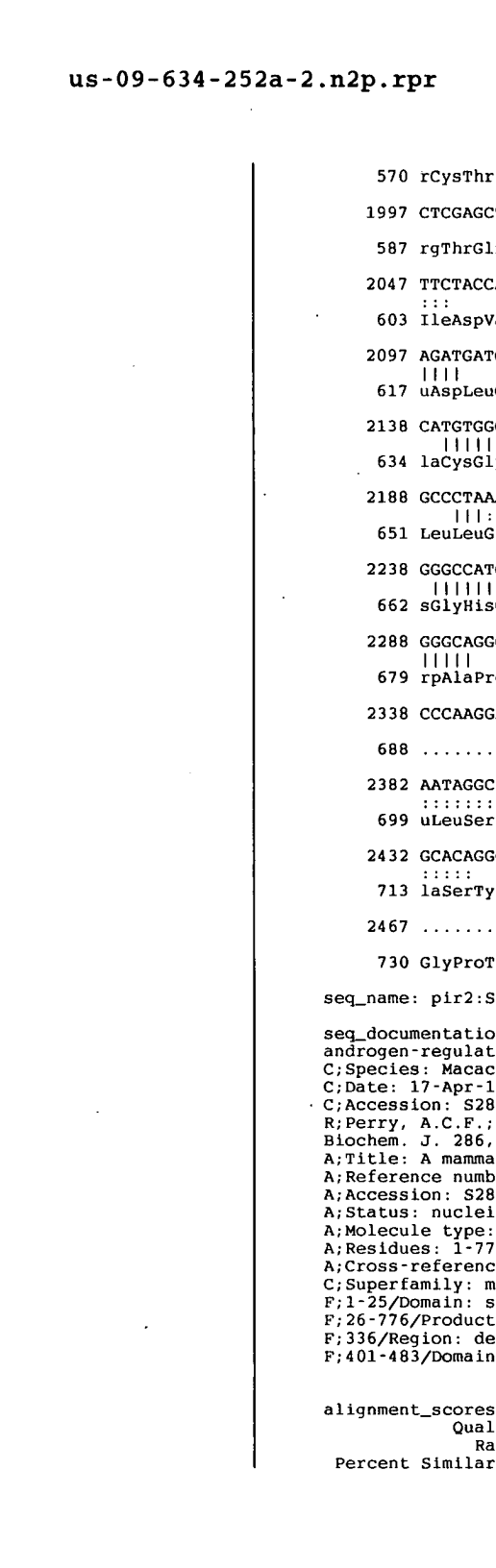
    951 TTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGG 1000
         :::      |||      |||   :::   :::|||:::   |||:::|
    231 .AspPheGlnHisLeuLeuAsnArgThrLeuGluValAlaLeuLeuLeuA 247

   1001 ATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTA 1050
        ||:::::::::::::      ||||||:::||||||:::||||||:::::
    247 spThrPhePheArg...ProLeuAsnValArgValAlaLeuValGlyLeu 262

   1051 GAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCA 1100
        |||:::|||||||:::::|||   ::::::|||:::   ||||||:::
    263 GluAlaTrpThrGlnArgAspLeuValGluIleSerProAsnProAlaVa 279

   1101 GATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATTAAGCAGCATGCT. 1149
```





seq_name: pir2:S28258

seq_documentation_block:
androgen-regulated epididymal protein precursor - crab-eating macaque
C;Species: Macaca fascicularis (crab-eating macaque)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 21-Jul-2000
C;Accession: S28258
R;Perry, A.C.F.; Jones, R.; Barker, P.J.; Hall, L.
Biochem. J. 286, 671-675, 1992
A;Title: A mammalian epididymal protein with remarkable sequence similarity to snake
A;Reference number: S28258; MUID:93038491
A;Accession: S28258
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-776 <PER>
A;Cross-references: EMBL:X66139; NID:g38062; PIDN:CAA46929.1; PID:g38063
C;Superfamily: mouse meltrin alpha; disintegrin homology
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-776/Product: androgen-regulated epididymal protein #status predicted <MAT>
F;336/Region: defective catalytic site
F;401-483/Domain: disintegrin homology <DIS>

alignment_scores:
Quality: 797.50 Length: 691
Ratio: 2.004 Gaps: 21
Percent Similarity: 57.598 Percent Identity: 29.522

```
alignment_block:
 US-09-634-252A-2 x S28258   ..

 Align seg 1/1  to: S28258  from: 1  to: 776

     457 TTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAA 506
         :::::::|||:::         :::   :::::::|||   |||:::   ::
          61 TyrGluIleLysLeuAsnArgLysThrLeuValLeuHisLeuLeuArgSe 77

     507 CAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAAATGGGA 556
         ::::   :::|||:::|||:::|||   |||      |||:::
          77 rArgGluPheLeuGlySerAsnTyrSerGluThrPheTyrSerMetLysG 94

     557 AACCACAGTACTCTAAGGGT.........GGAGAGCACTGTTACTACCAT 597
                 :::::::::              :::|||||||:::|||:::
          94 lyGluAlaPheThrArgHisLeuGlnIleMetAspHisCysPheTyrGln 110

     598 GGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAA 647
         |||||||||            ||||||   ::::::::::||||||||||
         111 GlySerIleValHisGluTyrAspSerAlaAlaSerIleSerThrCysAs 127

     648 TGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGC 697
         |||||||:::|||:::|||:::   :::         |||:::|||||||
         127 nGlyLeuArgGlyPhePheArgValAsnAspGlnArgTyrLeuIleGluP 144

     698 CACTAGAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCACATATAATC 747
         ||:::
         144 roVal..................................................... 145

     748 CAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTAT 797
                          :::||||||   :::   :::::::|||:::::
         146 .................LysTyrSerAspGluGlyGluHisLeuValPh 156

     798 GGAAAGAGGTGACCAGTGGCCCTTTCTC.............TCTGAAT 832
         ::::   :::   :::   |||:::::               :::||||
         156 eLysTyrAsnProArgValProTyrValAlaAsnTyrSerCysThrGluL 173

     833 TACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGT... 879
         ||::::::   ::::::         :::   :::   |||:::|||
         173 euAsnPheThrArgLysThrValProGlyAspThrGluSerGluGlyAsp 189

     880 ............ATATTTGAAGAAATGAAATATTTGGAACTTATGATTGT 917
                        :::::::::   ||||||:::||||||:::|||||
         190 ProLysMetLysAlaIleHisAsnGluLysTyrIleGluLeuPheIleVa 206

     918 TAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCA 967
         |   |||      :::|||:::::::::         |||         :
         206 lAlaAspAspThrValTyrArgArgAsnSerHisProHisAsnLysLeuA 223

     968 ACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAG 1017
         ::|||        ::::::||||||:::|||:::   |||||||||
         223 rgAsnArgIleTrpGlyMetValAsnPheValAsnMetIleTyrLys... 238

    1018 CAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAA 1067
            ||||||   :::|||:::|||||||:::::|||   ||||||:::::
         239 ThrLeuAsnIleHisValThrLeuValGlyIleGluIleTrpThrHisGl 255

    1068 GGATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCT 1117
         :|||:::|||::::::   :::|||            |||   :::||||
         255 uAspLysIleGluLeuHisSerAsnIleGluThrThrLeuLeuArgPheS 272

    1118 CAAAATACCGGCAGCGCATT.........AAGCAGCATGCTGATGCTGTG 1158
         ||:::::::::::|||||         ::::::      |||   |||
         272 erSerTrpGlnGluArgIleLeuLysThrArgLysAspPheAspHisVal 288

    1159 CACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTA 1208
            |||:::|||            ::::::          ||||||||||||||
         289 ValLeuLeuSerGlyLysTrpIleTyrThrHisValGlnGlyIleSerTy 305

    1209 CTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATG 1258
         |   :::|||:::|||            ::::::::::::   :::
         305 rProAlaGlyMetCysLeuProTyrTyrSerThrSerIleIleLysAspL 322

    1259 GTCTTCCA......ATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCT 1302
           ||||||            :::|||:::   :::::::::::::|||:::
         322 euLeuProAspThrAsnIleIleAlaAsnArgMetAlaHisGlnLeuGly 338

    1303 CAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATGTGA 1352
         :::|||||||||:::|||   :::                  :::
         339 HisAsnLeuGlyMetGlnHisAsp.................GluPhePr 349

    1353 CTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCC 1402
          ||||||         |||   |||:::|||::::::   |||
         349 oCysThrCysProSerGlyLysCysValMetAspSerAspGlySerIleP 366

    1403 ATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTA 1452
           :::   ||||||||||||||||||       :::|||:::::::::|||
         366 roAlaLeuLysPheSerLysCysSerGlnAsnGlnTyrHisGlnTyrLeu 382

    1453 CAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTA...TT 1499
         :::                :::|||:::::|||   |||            ||
         383 LysAspTyrLysProThrCysMetLeuAsnIleProPheProCysAsnPh 399

    1500 TGAGCCCACGGAA...TGTGGAAATGGATACGTGGAAGCTGGGGAGGAGT 1546
         |:::      :::   |||||||||      ::::::   ||||||||||
         399 eAspAspPheGlnPheCysGlyAsnLysLysLeuAspGluGlyGluGluC 416

    1547 GTGATTGTGGTTTTCATGTGGAATGCTATGGATTATGCTGT......AAG 1590
         ||||||||||          ||||||   :::   ||||||
         416 ysAspCysGlyProProGlnGluCysThrAsnProCysCysAspAlaHis 432

    1591 AAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAA 1640
           |||   |||:::   |||      |||::::::|||   ||||||::
         433 ThrCysValLeuLysProGlyPheThrCysAlaGluGlyGluCysCysGl 449

    1641 CAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTG 1690
         :      ||||||   ::::::      |||      ||||||   |||
         449 u......SerCysGlnIleLysLysAlaGlySerIleCysArgProAlaG 464

    1691 TGAACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGC 1740
           :::|||||||||:::   |||   ||||||||   |||      |||
         464 luAspGluCysAspPheProGluMetCysThrGlyHisSerProAlaCys 480

    1741 CCACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGG 1790
         |||   :::      :::   :::|||:::   |||:::::::::::||
         481 ProLysAspGlnPheArgValAsnGlyPheProCysLysAsnSerGluGl 497

    1791 CCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACA 1840
         |   |||:::   |||:::|||   ||||||   :::|||||||:::   :
         497 yTyrCysPheMetGlyLysCysProThrArgArgAspGlnCysSerGluL 514

    1841 TCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTG 1890
         :::::.      :::|||:::   |||      :::||||||   |||:::
         514 euPheAspAspGluAlaThrGluSerHisAspIleCysTyr...LysMet 529

    1891 AATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTG 1940
         ||||||:::|||::::::   |||   |||   :::      :::|||::
         530 AsnThrLysGlyAsnLysPheGlyTyrCysLysAsnLysGluAsnArgPh 546

    1941 GATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCA 1990
         ::::   |||::::::   ||||||   ||||||   :::   ||||||:
         546 eLeuProCysGluGluLysAspValArgCysGlyLysIleTyrCysThrG 563

    1991 ATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCA 2040
         ::                      |||:::|||:::      :::
         563 ly.....................GlyGluLeuSerTyrLeuLeuGlyGlu 572

    2041 ACTTCCTTCTACCAT..............CAAGGCCGGGTGATTGACTG 2075
           :::   ||||||              |||:::   ::::::   ||
         573 AspLysThrTyrHisLeuLysAspProGlnGlnAsnAlaThrValLysCy 589

    2076 CAGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTATGTAG 2125
         |:::   :::      :::      |||:::||||||:::|||   |||
         589 sLys...ThrIlePheLeuTyrHisAspSerThrAspIleGlyLeuValA 605
```

```
  2126 AAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGC 2175
       :::|||||||   ||||||   :::|||:::|||   :::   :::|||
   605 laSerGlyThrLysCysGlyAspGlyMetValCysAsnAsnGlyGluCys 621

  2176 CTACAAATTCAAGCCCTA...AATATGAGCAGCTGTCCACTCGATTCCAA 2222
       |||:::::::::   :::   |||   ::::::||||||   :::
   622 LeuAsnMetGluLysValTyrAsnSerThrAsnCysProSerGlnCysHi 638

  2223 GGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCA 2272
          :::      :::   |||||||:::                 |||
   638 sGluAsnProMetAspAspHisGlyLeu.............GlnCysH 650

  2273 TTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGTT 2322
       |||
   650 isCys..................................................... 651

   323 AGGAACCTTCACCCCCCCAAGGATGAAGGACCCAAGGGTCCTAGTGCC.. 2370
                            :::|||||||   :::::::|||   |||
   652 .....................GluGluGly...GlnAlaProValAlaTr 660

  2371 ....ACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAG 2416
               :::|||   :::   :::::::|||      |||:::|||   :
   660 pGluGluThrLeuAsnValThrAsnValAlaIleLeuIleValValLeuV 677

  2417 CTATTGTCCTTGGGGGCACAGGC 2439
       :::::|||:::   |||   |||
   677 alLeuValIleValGlyIleGly 684

seq_name: pir2:S55060

seq_documentation_block:
fertilin alpha-II - crab-eating macaque
C;Species: Macaca fascicularis (crab-eating macaque)
C;Date: 23-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 21-Jul-2000
C;Accession: S55060
R;Perry, A.C.F.; Gichuhi, P.M.; Jones, R.; Hall, L.
Biochem. J. 307, 843-850, 1995
A;Title: Cloning and analysis of monkey fertilin reveals novel alpha subunit isoforms.
A;Reference number: S55059; MUID:95260313
A;Accession: S55060
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-825 <PER>
A;Cross-references: EMBL:X79809; NID:g794074; PIDN:CAA56204.1; PID:g794075
C;Superfamily: mouse meltrin alpha; disintegrin homology
F;443-523/Domain: disintegrin homology <DIS>
F;[redacted]7/Active site: Glu #status predicted

alignment_scores:
           Quality:  779.00              Length:    730
             Ratio:  1.837                 Gaps:     28
 Percent Similarity:  58.082   Percent Identity: 29.315

alignment_block:
 US-09-634-252A-2 x S55060   ..

 Align seg 1/1  to: S55060  from: 1  to: 825

   295 AGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAATCACACTGCCTTC 344
       :::::::::   |||||||:::           ||||||::::::|||::
    73 AlaSerValTyrTyrSerSer.........TyrGluIleValIleProLy 86

   345 AAGACTCATATATTACATCAACCAAGACTCGGAAAGCCCTTATCACGTTC 394
       :|||||||:::         ::::::||||||
    86 sArgLeuMetValArgGlySerGluAspSer..................... 96

   395 TTGACACAAAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCTG 444
                                                      :::
    97 ..............................................Val 97

   445 GCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGA 494
          :::|||::::::   :::         |||:::|||   ::::::

    98 GluLysAlaThrTyrLeuLeuLeuMetGlnGlyGlnLysHisLeuValHi 114

   495 CCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACT 544
        |||   ::::::::::::   :::   :::::::::           |
   114 sLeuLysValLysArgSerHisPheValAsnAsnPheProValTyrSerT 131

   545 ACGAAAATGGG..............AAACCACAGTACTCTAAGGGTGGA 579
       ||:::||||||             :::|||    |||
   131 yrHisAsnGlyIleLeuGlyGlnGluSerProPheIleSer......... 144

   580 GAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGT 629
       :::   |||:::|||:::|||   |||:::||||||:::   |||   ||
   145 HisAspCysHisTyrGluGlyTyrIleGluGlyValSerGlySerPheVa 161

   630 GGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATGATACCT 679
       |:::::::::||||||   ||||||:::|||:::::::      :::
   161 lSerValAsnThrCysAlaGlyLeuArgGlyIleLeuIleLysGluGluL 178

   680 TCGTGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACA 729
            |||   |||||||||:::::                  |||:::
   178 ysSerTyrSerIleGluProMetAsp.................SerSer 188

   730 GGTCGACCACATATAATCCAGAAAACCTTGGCAGGACAG........... 768
          |||   :::   :::       |||:::|||   :::
   189 ArgArgPheGluHisValLeuTyrThrMetAlaHisGluAlaArgValSe 205

   769 ..............TATTCTAAGCAAATGAAGAATCTCACTATGGAAA 802
                          ::::::::::::   :::   :::   ::::
   205 rCysGlyValThrSerArgAspSerHisValValSerThrSerTrpGlnG 222

   803 GAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAAGG 852
       ::|||::::::   |||         :::||||||   |||:::
   222 lnGlySerArgLysPro......HisAspLeuGlnAlaLeuSerTyr... 235

   853 AAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATTT 902
                                 ::::::::::::   |||||||::
   236 .........................LeuTrpSerHisThrLysTyrVa 243

   903 GGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTT 952
       :|||::::::::|||   :::::::::   ::::::          |||
   243 lGluMetPheValValValAsnAsnGlnArgPheGlnMetTrpGlySer. 259

   953 CTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGAT 1002
                       |||:::   :::::::|||   ::::::||||||
   260 ................AsnIleAsnGluThrValGlnArgValValAsp 270

  1003 TCTATT..............TACAAGGAGCAGCTCAACACCAGGGTTGT 1037
          |||               :::   :::   :::|||||||:::||||
   271 ValIleAlaLeuAlaAsnGlyPheThrArgGlyIleAsnThrGluValVa 287

  1038 CCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCA 1087
       ||||::::::::|||   |||||||||   |||   ||||||:::::::
   287 lLeuAlaGlyMetGluIleTrpThrGluGlyAspLeuIleAspValAlaV 304

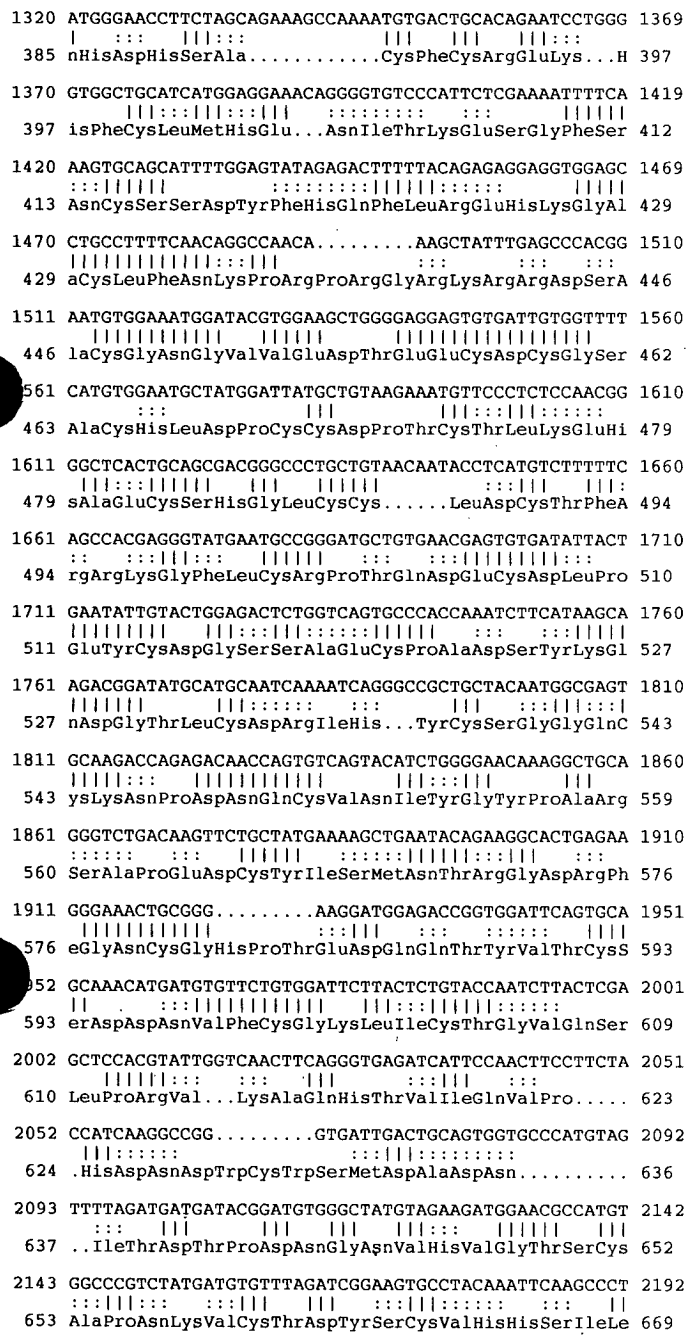
  1088 CCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATT 1137
       :::::             |||:::::::|||:::::::::||||||::::::
   304 alAspLeuGlnIleThrLeuArgAsnPheAsnArgTrpArgGlnGluMet 320

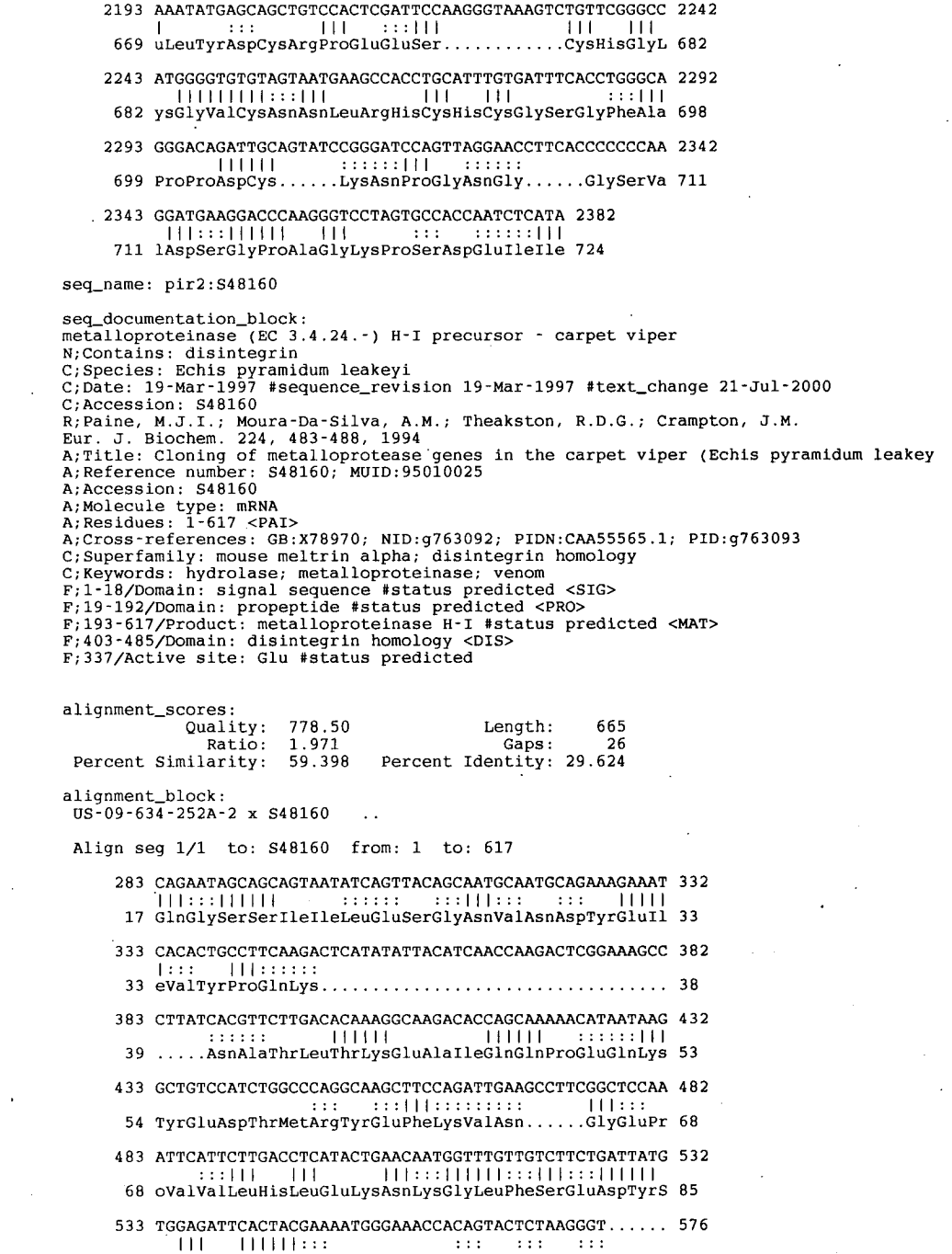
  1138 .........AAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGAC 1178
               :::         |||::::::|||:::|||      |||
   321 LeuPheArgArgAlaLysHisAspValAlaHisMetIle......ValGl 335

  1179 ATTTCAC...TATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTT 1225
          |||      ::::::::::::   ::::::::::::|||:::||||
   335 yHisHisProGlyGlnAsnThrGlyGlnAlaPheLeuSerGlyAlaCysS 352

  1226 CTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCA......ATG 1269
       ||         :::::::::|||:::::::::          :::
   352 erSerGlyPheAlaAlaAlaValGluSerPheHisHisGluAspMetLeu 368

  1270 GCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCA 1319
          |||   ::::::   ::::::|||:::::::||||||||||||||||
   369 LeuPheAlaAlaLeuMetValHisGluLeuGlyHisAsnLeuGlyIleGl 385
```





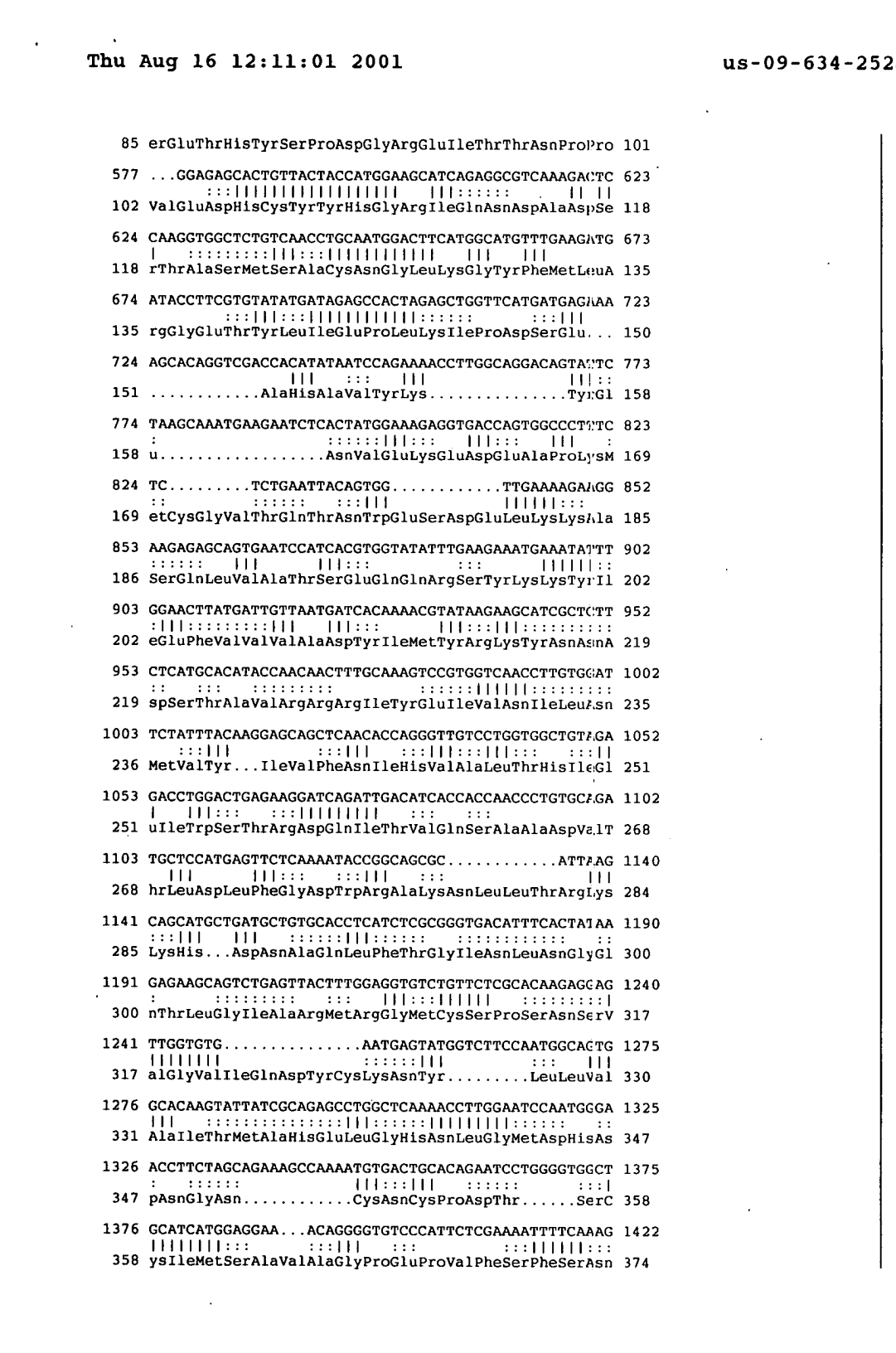
seq_name: pir2:S48160

seq_documentation_block:
metalloproteinase (EC 3.4.24.-) H-I precursor - carpet viper
N;Contains: disintegrin
C;Species: Echis pyramidum leakeyi
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 21-Jul-2000
C;Accession: S48160
R;Paine, M.J.I.; Moura-Da-Silva, A.M.; Theakston, R.D.G.; Crampton, J.M.
Eur. J. Biochem. 224, 483-488, 1994
A;Title: Cloning of metalloprotease genes in the carpet viper (Echis pyramidum leakey
A;Reference number: S48160; MUID:95010025
A;Accession: S48160
A;Molecule type: mRNA
A;Residues: 1-617 <PAI>
A;Cross-references: GB:X78970; NID:g763092; PIDN:CAA55565.1; PID:g763093
C;Superfamily: mouse meltrin alpha; disintegrin homology
C;Keywords: hydrolase; metalloproteinase; venom
F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-192/Domain: propeptide #status predicted <PRO>
F;193-617/Product: metalloproteinase H-I #status predicted <MAT>
F;403-485/Domain: disintegrin homology <DIS>
F;337/Active site: Glu #status predicted

alignment_scores:
Quality: 778.50 Length: 665
Ratio: 1.971 Gaps: 26
Percent Similarity: 59.398 Percent Identity: 29.624

alignment_block:
US-09-634-252A-2 x S48160 ..

Align seg 1/1 to: S48160 from: 1 to: 617



seq_name:·pir2:S28259

seq_documentation_block:
androgen-regulated epididymal protein precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 03-Dec-1999
C;Accession: S28259
R;Perry, A.C.F.; Jones, R.; Barker, P.J.; Hall, L.

```
Biochem. J. 286, 671-675, 1992
A;Title: A mammalian epididymal protein with remarkable sequence similarity to snake ven
A;Reference number: S28258; MUID:93038491
A;Accession: S28259
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-789 <PER>
A;Cross-references: EMBL:X66140; NID:g56069; PIDN:CAA46930.1; PID:g56070
C;Superfamily: mouse meltrin alpha; disintegrin homology
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-789/Product: androgen-regulated epididymal protein #status predicted <MAT>
F;336/Region: defective catalytic site
F;400-482/Domain: disintegrin homology <DIS>


alignment_scores:
            Quality:  775.00              Length:    747
              Ratio:  1.872                Gaps:     25
 Percent Similarity:  55.422   Percent Identity: 28.514

[redacted]nment_block:
09-634-252A-2 x S28259  ..

Align seg 1/1  to: S28259  from: 1  to: 789

     325 AAAGAAATCACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTC 374
         :::|||:::::   |||:::::|||            ::::::|||
      27 GlnGluLeuValHisProLysLysLeuSerLeuGluGlnLysArgAspLe 43

     375 GGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAAC 424
          |||      |||      ||||||      ::::::::::::
      43 uGluArg...IleHisAspSerAspThrProGluGluTyrGluGlu.... 57

     425 ATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTC 474
                             :::      ::::::|||:::
      58 ........................GluLeuLeuTyrGluIleLysLeuGly 66

     475 GGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTC 524
            :::   :::   |||   |||:::      :::   :::|||:::
      67 ArgLysThrLeuThrLeuHisLeuLeuLysAlaArgGluPheLeuAlaLe 83

     525 TGATTATGTGGAGATTCACTACGAAAATGGGAAACCACAGTACTCTAAGG 574
          :::|||   |||   :::|||:::      :::   :::   ::::::
      83 uAsnTyrSerGluThrTyrTyrAsnIleLysArgGluMetValThrArgH 100

     575 GTGGA.........GAGCACTGTTACTACCATGGAAGCATCAGAGGCGTC 615
                      :::||||||:::|||:::|||||||||
     100 isProGlnIleLeuAspHisCysPheTyrGlnGlySerIleIleHisGlu 116

     616 AAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTT 665
            ||||||   :::::::::||||||||||||||||||||:::|||:::||
     117 PheAspSerAlaAlaSerIleSerThrCysAsnGlyLeuArgGlyPhePh 133

     666 TGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATG 715
         |:::   :::         |||:::|||||||||:::
     133 eArgValAsnAspGlnArgTyrLeuIleGluProVal..... 145

     716 ATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCAGGA 765

     145 ..................................................... 145

     766 CAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAGGTGACCAGTG 815
         :::||||||   :::       :::|||:::::::::   :::   :::
     146 LysTyrSerAspGluGlyAspHisLeuValPheLysTyrAsnValLysAl 162

     816 GCCCTTTCTCTCT......GAATTACAGTGGTTGAAAAGAAGGAAGAGAG 859
          |||:::   :::       :::   :::   |||:::       |||::::
     162 aProTyrAlaThrAsnTyrSerCysGluGlyLeuAsnPheThrLysLysS 179

     860 CAGTGAATCCATCACGTGGTATATTTGAAGAA..... 891
         :::::       :::       |||:::||||||
     179 erThrLeuIleAspAlaLysIleIleGluGluHisLysValGluAspTyr 195

     892 ......ATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTA 935
                 |||:::::::||||||:::::|||   |||:::   :::||
     196 HisLysGluLysPheIleGluLeuPheValValAlaAspGluPheValTy 212

     936 TAAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCG 985
         |:::::::::   :::   :::          ::::::          ::::
     212 rArgArgAsnSerLysProGlnAsnLysLeuArgLysArgIleTrpGlyM 229

     986 TGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTT 1035
         ::||||||:::|||:::   |||||||||       ||||||   ||||||
     229 etValAsnPheValAsnMetIleTyrLys...AlaLeuAsnIleArgVal 244

    1036 GTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCAC 1085
         :::|||:::::::::|||   |||:::      |||:::|||:::|||::
     245 ThrLeuThrGlyMetGluIleTrpSerAlaGlyAspGluIleGluIleVa 261

    1086 CACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGCA 1135
         ::::|||      :::   |||   :::||||||   :::::::::   :
     261 lSerAsnLeuGluSerThrLeuLeuHisPheSerThrTrpGlnGluThrV 278

    1136 TT.........AAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTG 1176
         ::          ::::::      |||   |||   |||:::|||
     278 alLeuLysLysArgLysAspPheAspHisValIleLeuLeuSerGlyLys 294

    1177 ACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTC 1226
            ::::::   :::   :::::::::::::|||   ||||||:::||||::
     295 TrpLeuTyrThrSerMetGlnGlyIleAlaTyrProGlyGlyIleCysGl 311

    1227 TCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCA......ATGG 1270
         :      |||:::   :::|||   :::      ||||||
     311 nThrLeuArgSerCysSerValValLysAspLeuLeuProAspValAsnI 328

    1271 CAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAA 1320
           :::::::::   ::::::::::::|||:::::::::||||||:::
     328 leIleGlyAsnArgMetAlaHisGlnLeuGlyHisSerLeuGlyMet... 343

    1321 TGGGAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGG 1370
                        |||              ||||||          ||
     344 ..............ArgHisAspAspPheProCysThrCysProLeuGl 355

    1371 TGGCTGCATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAA 1420
         |   |||:::|||       :::|||          :::   ||||||||||
     355 yLysCysValMet...GlyAlaGlySerIleProAlaIleLysPheSerL 371

    1421 AGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCC 1470
         ||||||||       :::|||::::::||||||::::::          |||
     371 ysCysSerGlnThrGlnTyrGlnGlnPheLeuLysAsnGlnLysProAla 387

    1471 TGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAA........ 1512
         |||:::::|||:::|||                ||| |||
     388 CysIleLeuAsnAsnProLeu.............ProGluGluPheAsnAs 400

    1513 ..........TGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATT 1552
                   |||||||||       |||:::   ||||||||||||||||
     400 pTyrProPheCysGlyAsnLysLysValAspGluGlyGluGluCysAspC 417

    1553 GTGGTTTTCATGTGGAATGCTATGGATTATGCTGT......AAGAAATGT 1596
         |||||       ||||||   :::   ||||||          ||||||
     417 ysGlyProValGlnGluCysThrAsnProCysCysAspAlaHisLysCys 433

    1597 TCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATAC 1646
            |||:::   |||       |||   :::|||   |||||||:::
     434 ValLeuLysProGlyPheThrCysValGluGlyGluCysCysGlu..... 448

    1647 CTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACG 1696
          ||||||   ::::::   :::|||       ||||||   |||   ||||
     449 .SerCysGlnMetLysLysGluGlyValIleCysArgProAlaLysAsnG 465

    1697 AGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCA 1746
         ||||||||||:::|||   |||||||||   |||   :::||||||
     465 luCysAspIleSerGluValCysThrGlyTyrSerProGluCysProLys 481

    1747 AATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTG 1796
```

```
     :::      :::   :::|||:::   |||:::::::::::::|||   ||
  482 AspGluSerGlnAlaAsnGlyPheProCysLysAsnGlyGluGlyTyrCy 498

 1797 CTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGG 1846
      |:::   |||   |||   |||||||||:::||||||      :::::::
  498 sPheMetGlyLeuCysProThrArgAspAspGlnCysAlaGluLeuPheS 515

 1847 GAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACA 1896
      ::      |||      |||   ::::::||||||   ::::::|||
  515 erGlyGlyAlaGluGluSerHisSerLeuCysTyr...ArgMetAsnGln 530

 1897 GAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCA 1946
      :::|||::::::   |||   |||   :::      :::   ::::::
  531 LysGlyAsnArgPheGlyTyrCysLysAsnLysAspAsnThrPheValPr 547

  947 GTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTA 1996
      |||::::::   |||:::   ||||||   :::   |||||||:::
  547 oCysGluGluLysAspLeuLysCysGlyLysIleTyrCysThrGlyGlyA 564

 1997 CTCGAGCTCCACGTATTGGTCAA.....CTT 2022
        |||:::   ::::::|||:::      :::
  564 rgArgSerAlaHisLeuGlyGluAspLysThrTyrAsnLeuLysAsnVal 580

 2023 CAGGGTGAGATC.....ATTCCAACTTCCTTC...TACCATCAAGG 2060
      :::   :::|||            |||   |||   ||||||:::::
  581 LysGlnAsnIleSerIleLysCysLysThrMetPheLeuTyrHisAsnSe 597

 2061 CCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACGG 2110
      :|||                                          |
  597 rArg.....A 599

 2111 ATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGT 2160
      ||:::|||   |||::::::||||||   ||||||   :::|||:::|||
  599 spMetGlyLeuValAsnSerGlyThrLysCysGlyGluGlyMetValCys 615

 2161 TTAGATCGGAAGTGCCTACAAATTCAA...GCCCTAAATATGAGCAGCTG 2207
         :::   :::|||:::::::::::   |||   |||   :::   ||
  616 SerAsnGlyGluCysIleGluMetGluLysAlaTyrAsnSerThrIleCy 632

 2208 TCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTA 2257
      |              ::::::::::|||:::   :::   |||   :::|
  632 s.....SerSerLeuCysAspGluAsnAspValAspAspA 644

 2258 ATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGT 2307
      |||||      |||   |||
  644 snGluProAspCysGlnCys..... 550

  308 ATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAAGGATGAAGGACCC.. 2355
                                          :::|||||||||
  651 .....GluGluGlyProIl 655

 2356 .....AAGGGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCTG 2398
           :::      :::|||   ||||||         ||||||
  655 eIleThrGluTrpGlyGluAlaLeuAsnLeuThrSerValSerIleMetV 672

 2399 GTGCCATCCTGGTAGCAGCTATTGTCCTTGGGGGCACAGGC 2439
        ::::::||||||   :::|||:::      |||:::|||
  672 alValValLeuValMetValIleIle......GlyValGly 683

seq_name: pir2:S55059

seq_documentation_block:
fertilin alpha-I - crab-eating macaque
C;Species: Macaca fascicularis (crab-eating macaque)
C;Date: 23-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 21-Jul-2000
C;Accession: S55059
R;Perry, A.C.F.; Gichuhi, P.M.; Jones, R.; Hall, L.
Biochem. J. 307, 843-850, 1995
A;Title: Cloning and analysis of monkey fertilin reveals novel alpha subunit isoforms.
A;Reference number: S55059; MUID:95260313
A;Accession: S55059
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-905 <PER>
A;Cross-references: EMBL:X79808; NID:g794072; PIDN:CAA56203.1; PID:g794073
C;Superfamily: mouse meltrin alpha; disintegrin homology
F;442-522/Domain: disintegrin homology <DIS>
F;376/Active site: Glu #status predicted


alignment_scores:
            Quality:  775.00            Length:    769
              Ratio:  1.757               Gaps:     34
 Percent Similarity:  57.347   Percent Identity: 29.649

alignment_block:
 US-09-634-252A-2 x S55059   ..

 Align seg 1/1  to: S55059  from: 1  to: 905

  227 ATGAAACTGCAGAAAAAAATTTGG......GAGTCCTGGCAGATGAAGAC 270
      :::|||::::::   :::   |||      :::::::|||:::::::::
   27 ValLysIleLysPheGlnThrTrpAlaProGlnLysTrpAsnLeuArgMe 43

  271 A.............ATACATTGCAACAGA.....ATAGCAG 293
                              |||   |||        :::
   43 tGlyLeuValProGlyProSerCysIleArgLeuGluIleLeuLeuLeuL 60

  294 CAGTAATATCAGTTACAGCAATGCAATGCAGAAA..... 327
        ||||||   |||   :::|||:::|||:::
   60 euValIlePheValProSerMetHisCysHisLeuGlySerIleTyrTyr 76

  328 .........GAAATCACACTGCCTTCAAGACTCATATATTACATCAACCA 368
              ||||||   :::|||:::||||||                ||
   77 SerPheTyrGluIleIleIleProLysArgLeuThrVal.........Gl 90

  369 AGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAG. 417
      |   ::::::|||      :::   |||:::                 |||
   90 nGlyGlyAspSerSerValGluGlyLeuSerTyrLeuLeuPheMetGlnG 107

  418 ..CAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATT 465
        |||||||||         |||||||||
  107 lyGlnLysHis......LeuValHisLeu..... 114

  466 GAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTG.....AA 506
      ::::::      |||   |||:::   ::::::   :::        ::
  115 LysValLysArgSerHisPheValAsnAsnPheProValTyrSerTyrHi 131

  507 CAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAAATGGGA 556
      :||||||:::|||:::::::::         |||   ::::::
  131 sAsnGlyIleLeuGlyGlnGluSerProPheIleSerHisAsp..... 145

  557 AACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATC 606
                                    |||:::|||:::|||   |||
  146 ............................CysHisTyrGluGlyTyrIle 152

  607 AGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCA 656
      :::||||||:::   |||   |||::::::::::||||||   ||||||::
  153 GluGlyValSerGlySerPheValSerValAsnThrCysAlaGlyLeuAr 169

  657 TGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGC 706
      :|||::::::      :::         |||   |||||||||::::::
  169 gGlyIleLeuIleLysGluGluLysSerTyrSerIleGluProMetAsp. 185

  707 TGGTTCATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAACC 756
                        |||:::   |||   :::   :::      |||
  186 ................SerSerArgArgPheGluHisValLeuTyrThr 196

  757 TTGGCAGGACAG.........................TATTCTAAGCA 779
      :::|||   :::                        :::::::
  197 MetAlaHisGluAlaArgValSerCysGlyValThrSerArgAspSerHi 213

  780 AATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTG 829
      ::::   :::   :::   ::::::|||::::::   |||        :
  213 sValValSerThrSerTrpGlnGlnGlySerArgLysPro......HisA 228
```

```
 830 AATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGT 879
     ::|||||   |||:::
 228 spLeuGlnAlaLeuSerTyr............................. 234

 880 ATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAA 929
     ::::::::::::   ||||||:::|||:::::::::|||   ::::::::
 235 LeuTrpSerHisThrLysTyrValGluMetPheValValValAsnAsnGl 251

 930 AACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAA 979
     :   ::::::          |||                  |||:::   :
 251 nArgPheGlnMetTrpGlySer..................AsnIleAsnG 262

 980 AGTCCGTGGTCAACCTTGTGGATTCTATT..............TACAAG 1014
     :::::|||   ::::::||||||   |||              :::
 262 luThrValGlnArgValValAspValIleAlaLeuAlaAsnGlyPheThr 278

1015 GAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGA 1064
     :::   :::||||||:::|||||||||:::::::::|||   ||||||||
 279 ArgGlyIleAsnThrGluValValLeuAlaGlyMetGluIleTrpThrGl 295

 065 GAAGGATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGT 1114
     |   |||   ||||||::::::::::::          |||::::::|
 295 uGlyAspLeuIleAspValAlaValAspLeuGlnIleThrLeuArgAsnP 312

1115 TCTCAAAATACCGGCAGCGCATT.........AAGCAGCATGCTGATGCT 1155
     ||:::::::::||||||:::::         :::         |||:::
 312 heAsnArgTrpArgGlnGluMetLeuPheArgArgAlaLysHisAspVal 328

1156 GTGCACCTCATCTCGCGGGTGACATTTCAC...TATAAGAGAAGCAGTCT 1202
     :::|||:::|||      |||      |||      ::::::::::::
 329 AlaHisMetIle......ValGlyHisHisProGlyGlnAsnThrGlyGl 343

1203 GAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATG 1252
      ::::::::::::|||:::||||||         :::::::::|||::::
 343 nAlaPheLeuSerGlyAlaCysSerSerGlyPheAlaAlaAlaValGluS 360

1253 AGTATGGTCTTCCA......ATGGCAGTGGCACAAGTATTATCGCAGAGC 1296
     :::::             :::      |||   ::::::   ::::::
 360 erPheHisHisGluAspMetLeuLeuPheAlaAlaLeuMetValHisGlu 376

1297 CTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAA 1346
     |||::::::||||||||||||||   :::   |||:::
 377 LeuGlyHisAsnLeuGlyIleGlnHisAspHisSerAla........... 389

1347 ATGTGACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGG 1396
      |||   |||   |||:::        |||:::|||:::|||   ::::
 390 .CysPheCysArgGluLys...HisPheCysLeuMetHisGlu...AsnI 404

1397 TGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGAC 1446
     :::::   :::      ||||||:::||||||         :::::::::
 404 leThrLysGluSerGlyPheSerAsnCysSerSerAspTyrPheHisGln 420

1447 TTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACA..... 1491
     ||||||::::::      ||||||||||||||||||:::|||
 421 PheLeuArgGluHisLysGlyAlaCysLeuPheAsnLysProArgProAr 437

1492 ....AAGCTATTTGAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTG 1537
         :::      :::   :::   ||||||||||||   ||||||
 437 gGlyArgLysArgArgAspSerAlaCysGlyAsnGlyValValGluAspT 454

1538 GGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTATGGATTATGCTGT 1587
       ||||||||||||||||         :::             |||
 454 hrGluGluCysAspCysGlySerAlaCysHisLeuAspProCysCysAsp 470

1588 AAGAAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTG 1637
           |||:::|||::::::   |||:::||||||   |||   |||||
 471 ProThrCysThrLeuLysGluHisAlaGluCysSerHisGlyLeuCysCy 487

1638 TAACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATG 1687
     |         :::|||   |||:::   :::|||:::   ||||||   :
 487 s......LeuAspCysThrPheArgArgLysGlyPheLeuCysArgProT 502

1688 CTGTGAACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAG 1737
```

```
         ::   :::||||||||||:::   ||||||||   |||:::|||::::::
 502 hrGlnAspGluCysAspLeuProGluTyrCysAspGlySerSerAlaGlu 518

1738 TGCCCACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCA 1787
     ||||||   :::   :::||||||||||||      |||::::::   ::
 519 CysProAlaAspSerTyrLysGlnAspGlyThrLeuCysAspArgIleHi 535

1788 GGGCCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGT 1837
     :      |||   :::|||:::||||||:::   ||||||||||||
 535 s...TyrCysSerGlyGlyGlnCysLysAsnProAspAsnGlnCysValA 551

1838 ACATCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAG 1887
      |||:::|||      |||   ::::::  :::   ||||||   :::
 551 snIleTyrGlyTyrProAlaArgSerAlaProGluAspCysTyrIleSer 567

1888 CTGAATACAGAAGGCACTGAGAAGGGAAACTGCGGG.........AAGGA 1928
     :::||||||:::|||   :::   ||||||||||||         :::||
 568 MetAsnThrArgGlyAspArgPheGlyAsnCysGlyHisProThrGluAs 584

1929 TGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCT 1978
     |   :::   ::::::   ||||||      :::||||||||||||   |
 584 pGlnGlnThrTyrValThrCysSerAspAspAsnValPheCysGlyLysL 601

1979 TACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGT 2028
     ||:::||||||::::::          ||||||:::   :::   |||
 601 euIleCysThrGlyValGlnSerLeuProArgVal...LysAlaGlnHis 616

2029 GAGATCATTCCAACTTCCTTCTACCATCAAGGCCGG.........GTGAT 2069
        :::|||   :::          |||::::::          ::
 617 ThrValIleGlnValPro......HisAspAsnAspTrpCysTrpSerMe 631

2070 TGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCT 2119
     :|||:::::::::            :::   |||      |||   |||
 631 tAspAlaAspAsn............IleThrAspThrProAspAsnGlyA 644

2120 ATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGG 2169
       |||:::   ||||||   |||:::|||:::   :::|||   |||
 644 snValHisValGlyThrSerCysAlaProAsnLysValCysThrAspTyr 660

2170 AAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTC 2219
     :::|||::::::   :::   |||      :::      |||   :::||
 661 SerCysValHisHisSerIleLeuLeuTyrAspCysArgProGluGluSe 677

2220 CAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCT 2269
     |            |||   |||   |||||||||:::|||         |
 677 r............CysHisGlyLysGlyValCysAsnAsnLeuArgHisC 690

2270 GCATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCA 2319
     ||   |||:::      :::|||      ||||||      ::::::|||
 690 ysHisCysGluSerGlyPheAlaProProAspCys......LysAsnPro 704

2320 GTTAGGAACCTTCACCCCCCCAAGGATGAAGGACCCAAGGGTCCTAGTGC 2369
        ::::::              |||:::|||||   |||   :::::
 705 GlyAsnGly......GlySerValAspSerGlyProProGlyMetGlnVa 719

2370 CACCAAT 2376
     :||||||
 719 lThrAsn 721
```

seq_name: pir2:S71949

seq_documentation_block:
metalloproteinase 12 (EC 3.4.24.-) precursor - human
N;Alternate names: disintegrin 12; myeloma cell metalloproteinase MCMP
C;Species: Homo sapiens (man)
C;Date: 19-Mar-1998 #sequence_revision 17-Apr-1998 #text_change 04-Feb-2000
C;Accession: S71949; PC4264
R;McKie, N.; Dallas, D.J.; Edwards, T.; Apperley, J.F.; Russell, R.G.G.; Croucher, P.
Biochem. J. 318, 459-462, 1996
A;Title: Cloning of a novel membrane-linked metalloproteinase from human myeloma cell
A;Reference number: S71949; MUID:96404892
A;Accession: S71949
A;Status: nucleic acid sequence not shown; not compared with conceptual translation

```
A;Molecule type: mRNA
A;Residues: 1-660 <MCK>
A;Experimental source: myeloma cells
R;McKie, N.; Edwards, T.; Dallas, D.J.; Houghton, A.; Stringer, B.; Graham, R.; Russell,
Biochem. Biophys. Res. Commun. 230, 335-339, 1997
A;Title: Expression of members of a novel membrane linked metalloproteinase family (ADAM
A;Reference number: PC4263; MUID:97168971
A;Accession: PC4264
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-337 <MCF>
A;Experimental source: articular chondrocyte
C;Function:
A;Description: involved in cell/cell and cell/matrix interaction in a variety of cell ty
[illegible]ote: membrane-bound; belongs to reprolysin family of metalloproteinases
[illegible]perfamily: mouse meltrin alpha; disintegrin homology
[illegible]words: hydrolase; metalloproteinase; zinc
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-660/Product: pre-metalloproteinase 12 #status predicted <MAT>
F;54-59/Region: autoinhibitory
F;295-378/Domain: disintegrin homology <DIS>
F;574-598/Region: epidermal growth factor-like
F;622-642/Domain: transmembrane #status predicted <TM1>
F;57,229,233,239/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status
F;229,233,239/Binding site: zinc, catalytic (His) (active) #status predicted
F;230/Active site: Glu #status predicted


alignment_scores:
            Quality:  774.00              Length:    622
              Ratio:  2.109                Gaps:     20
 Percent Similarity:  59.003   Percent Identity: 30.707

alignment_block:
 US-09-634-252A-2 x S71949   ..

 Align seg 1/1  to: S71949  from: 1  to: 660

     634 CTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATGATACCTTCGT 683
         |||:::|||         |||:::|||:::::::::   ::::::
      12 LeuAlaThrValLeuAspLeuArgGlyLeuLeuHisLeuGluAsnAlaSe 28

     684 GTATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACAGGTC 733
          |||   |||||||||||||:::::                 |||:::
      28 rTyrGlyIleGluProLeuGlnMet..............SerSerHisP 40

     734 GACCACATATAATC.......................CAGAAAACCTGG 759
              |||||||||                       :::
      40 heGluHisIleIleTyrArgMetAspAspValTyrLysGluProLeuLys 56

     760 GCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAGGTGA 809
         :::|||      :::|||:::::::::::   |||   :::      ::
      57 CysGlyValSerAsnLysAspIleGluLysGluThrAlaLysAspGluGl 73

     810 CCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAAGAGAG 859
         ::::   |||         ::::::      |||   ||||||:::|| |
      73 uGluGluPro......ProSerMetThrGlnLeuLeuArgArgArgArgA 88

     860 CAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACCT 909
         |||||   |||                 :::   :::|||:::|||| |
      88 laValLeuPro.................GlnThrArgTyrValGluLeu 98

     910 ATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGC 959
         :::|||||||   |||   :::   |||               :::::::
      99 PheIleValValAspLysGluArgTyrAspMetMetGlyArgAsnGlnTh 115

     960 ACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTT 1009
         :   :::::::::          :::::|||   :::||||||:::|
     115 rAlaValArgGluGluMetIleLeuLeuAlaAsnTyrLeuAspSerMetT 132

    1010 ACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGG 1059
         ||      :::||||||   |||:::|||||||||:::::::|||   |||
     132 yr...IleMetLeuAsnIleArgIleValLeuValGlyLeuGluIleTrp 147

    1060 ACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCA 1109
         |||:::   :::   |||:::|||:::   :::      :::::|||
     148 ThrAsnGlyAsnLeuIleAsnIleValGlyGlyAlaGlyAspValLeuGl 164

    1110 TGAGTTCTCAAAATACCGGCAGCGCATT.........AAGCAGCATGCTG 1150
          :::|||   ::::::|||:::::::::         :::::::::   |
     164 yAsnPheValGlnTrpArgGluLysPheLeuIleThrArgArgArgHisA 181

    1151 ATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGT 1200
         ||:::::::::|||:::   :::      |||                 ::::::
     181 spSerAlaGlnLeuValLeuLysLysGlyPheGly...GlyThrAlaGly 196

    1201 CTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAA 1250
         :::::::::   |||   |||||||||||||:::::::::   |||:::||
     197 MetAlaPheValGlyThrValCysSerArgSerHisAlaGlyGlyIleAs 213

    1251 TGAGTATGGT......CTTCCAATGGCAGTGGCACAAGTATTATCGCAGA 1294
         |   :::|||               :::   |||:::::::::::::::::
     213 nValPheGlyGlnIleThrValGluThrPheAlaSerIleValAlaHisG 230

    1295 GCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCA 1344
         ::|||::::::|||||||||::::::   :::   :::
     230 luLeuGlyHisAsnLeuGlyMetAsnHisAspAspGlyArg......... 243

    1345 AAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGG 1394
               |||||||:::   :::      :::||||||||||   :::::::||
     244 ......AspCysSerCysGlyAlaLysSerCysIleMet...AsnSerGl 257

    1395 GGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAG 1444
         |:::|||   ||||||:::||||||:::|||||      :::::::::
     257 yAlaSerGlySerArgAsnPheSerSerCysSerAlaGluAspPheGluL 274

    1445 ACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAAC......AGGCCA 1488
           :::      :::   ||||||   ||||||:::|||      :::|||
     274 ysLeuThrLeuAsnLysGlyGlyAsnCysLeuLeuAsnIleProLysPro 290

    1489 ACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGG 1538
            :::   ::::::      :::|||||||||      |||:::|||||
     291 AspGluAlaTyrSerAlaProSerCysGlyAsnLysLeuValAspAlaGl 307

    1539 GGAGGAGTGTGATTGTGGTTTTCATGTGGAATGC...TATGGATTATGCT 1585
         ||||||||||||||||||         ||||||             ||||
     307 yGluGluCysAspCysGlyThrProLysGluCysGluLeuAspProCysC 324

    1586 GTAAG......AAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGG 1629
         ||:::         |||:::|||::::::   |||:::|||:::   |||
     324 ysGluGlySerThrCysLysLeuLysSerPheAlaGluCysAlaTyrGly 340

    1630 CCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATG 1679
            ||||||:::::       |||   |||   |||   |||      ||
     341 AspCysCysLysAsp......CysArgPheLeuProGlyGlyThrLeuCy 355

    1680 CCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTACTGGAGACT 1729
         ||||      ::::::||||||||::   |||||||||:::|||:::|
     355 sArgGlyLysThrSerGluCysAspValProGluTyrCysAsnGlySerS 372

    1730 CTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATGCATGCAAT 1779
         ||      |||   |||::::::      |||:::||||||   |||:::
     372 erGlnPheCysGlnProAspValPheIleGlnAsnGlyTyrProCysGln 388

    1780 CAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAACCA 1829
         :::|||::::::   ||||||||||||   |||:::      |||   ||
     389 AsnAsnLysAlaTyrCysTyrAsnGlyMetCysGlnTyrTyrAspAlaGl 405

    1830 GTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCT 1879
         |||||||   |||:::|||:::||||||   ::::::   |||   |||:
     405 nCysGlnValIlePheGlySerLysAlaLysAlaAlaProLysAspCysP 422

    1880 ATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGAT 1929
         ::   ::::::|||::::::|||   :::   ||||||||||||   :::
     422 heIleGluValAsnSerLysGlyAspArgPheGlyAsnCysGlyPheSer 438
```

```
1930 GGAGACCGGTGG..................................... 1941
     |||:::::::::
 439 GlyAsnGluTyrLysLysCysAlaThrGlyLeuSerLeuLysPheHisAl 455

1941 ..................................................... 1941

 455 aProPheLeuSerThrMetLeuGlnGluAlaValArgGlnThrGlyThrT 472

1942 ..............ATTCAGTGCAGCAAACATGAT.............. 1962
                    :::   |||   |||   |||
 472 yrLeuGlyGlySerValCysCysMetLysSerAspCysArgIleValThr 488

1963 ............GTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCG 2000
                 ::::::||||||   |||   |||   |||:::   ::
 489 LeuValLysAsnAlaLeuCysGlyLysLeuGlnCysGluAsnValGlnGl 505

2001 AGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCT 2050
     :   |||   :::|||                |||:::|||:::   :::
 505 uIleProValPheGly..............IleValProAlaIleIleG 517

2051 ACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGAT 2100
           :::|||         |||   |||:::        |||
 517 lnThrProSerArgGlyThrLysCysTrpGlyValAspPheGlnLeuGly 533

2101 GATGATACG...GATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCC 2147
     :::|||:::   |||   |||   |||:::::|||||   ||||||
 534 SerAspValProAspProGlyMetValAsnGluGlyThrLysCysGlyAl 550

2148 GTCTATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATA 2197
      :::   :::|||   :::   :::|||:::::::   ::::::||||||
 550 aGlyLysIleCysArgAsnPheGlnCysValAspAlaSerValLeuAsnT 567

2198 TGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGG 2247
           :::|||   :::::::::|||         |||   |||||||||
 567 yr...AspCysAspValGlnLysLys.........CysHisGlyHisGly 579

2248 GTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGAC 2297
     ||||||:::::::::   :::|||   |||:::      ||||||
 580 ValCysAsnSerAsnLysAsnCysHisCysGluAsnGlyTrpAlaProPr 596

2298 AGATTGCAGTATCCGG 2313
      :::|||:::   :::
 596 oAsnCysGluThrLys 601
```

seq_name: pir2:G02937

seq_documentation_block:
fertilin beta - crab-eating macaque
C;Species: Macaca fascicularis (crab-eating macaque)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 01-Dec-2000
C;Accession: GQ2937; S55061
[redacted]marao, C.S.; Myles, D.G.; White, J.M.; Primakoff, P.
[redacted]tted to the EMBL Data Library, August 1995
[redacted]erence number: G12615
[redacted]cession: G02937
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-735 <RAM>
A;Cross-references: EMBL:U33959; NID:g998339; PID:g998340
R;Perry, A.C.F.; Gichuhi, P.M.; Jones, R.; Hall, L.
Biochem. J. 307, 843-850, 1995
A;Title: Cloning and analysis of monkey fertilin reveals novel alpha subunit isoforms.
A;Reference number: S55059; MUID:95260313
A;Accession: S55061
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-722,'S',724-735 <PER>
A;Cross-references: EMBL:X77653; NID:g794076; PIDN:CAA54733.1; PID:g794077
C;Superfamily: mouse meltrin alpha; disintegrin homology
F;383-468/Domain: disintegrin homology <DIS>

alignment_scores:

```
          Quality:  754.50          Length:    714
            Ratio:  1.858            Gaps:     26
Percent Similarity:  56.863   Percent Identity: 27.871

alignment_block:
US-09-634-252A-2 x G02937   ..

Align seg 1/1  to: G02937  from: 1  to: 735

 313 AATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACAT 362
     :::::::::      :::||||||:::|||:::::::::         ||
  21 AspSerLeuProValGlnIleThrValProGluLysIleArgSerIleIl 37

 363 CAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGAC 412
     |:::::::::   ||||||
  37 eLysGluGluIleGluSer..................................... 43

 413 ACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAG 462
                                          |||:::|||::::::
  44 .....................................GlnValSerTyrLys 48

 463 ATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGG 512
     |||         |||:::   :::      :::|||:::   :::|||
  49 IleValIleGluGlyLysProTyrThrAlaAsnLeuMetGlnLysAsnPh 65

 513 TTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAAATGGG....... 555
      |||       |||        :::::::::      ||||||
  65 eLeu......SerHisAsnPheArgValTyrSerTyrAsnGlyThrGlyI 80

 556 .....AAACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGA 600
          ||||||      :::         :::   |||:::|||:::|||
  80 leMetLysProLeuAspGlnAspPheGlnAsnPheCysHisTyrGlnGly 96

 601 AGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGG 650
      |||:::|||         |||   :::   :::|||||||||:::|||
  97 TyrIleGluGlyTyrProLysSerValAlaMetValSerThrCysThrGl 113

 651 ACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCAC 700
     ||||:::|||::::::::::   ::::::      |||  ||||||||||
 113 yLeuArgGlyLeuLeuGlnPheGluAsnValSerTyrGlyIleGluProL 130

 701 TAGAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCACATATAATCCAG 750
     ||              |||:::|||:::|||      |||:::|||
 130 eu..............GluSerSerValGlyPheGluHisValIleTyr 141

 751 AAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGA 800
     ::::::         ::::::::::|||         :::|||      ||
 142 GlnVal.........LysHisLysLysAlaAspValSerLeuTyrAsnGl 155

 801 AAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAA 850
     |:::                        |||   :::   |||:::   :
 155 uLysAsp.................IleGluSerArgAspLeuSerPheL 166

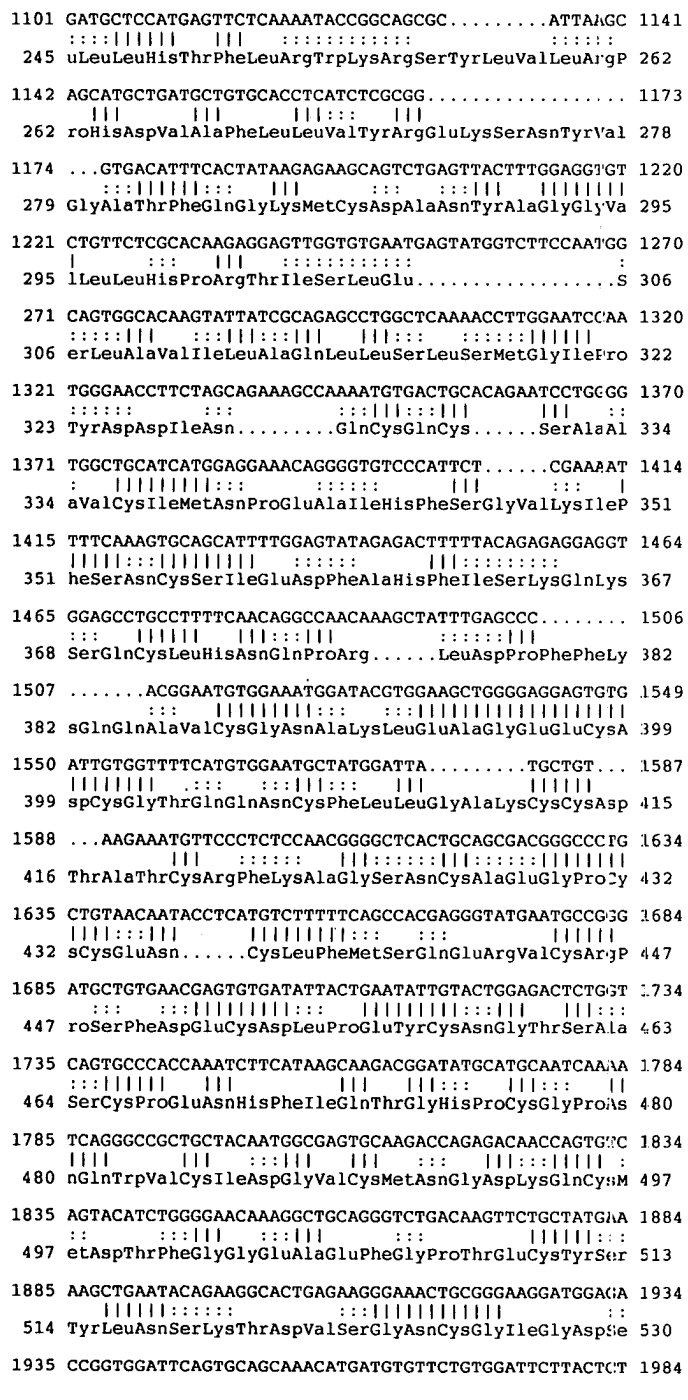
 851 GGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATAT 900
     ::   ::::::::::::|||::::::         :::      ||||||
 166 ysLeuGlnSerIleGluProGlnLys.........AspPheAlaLysTyr 179

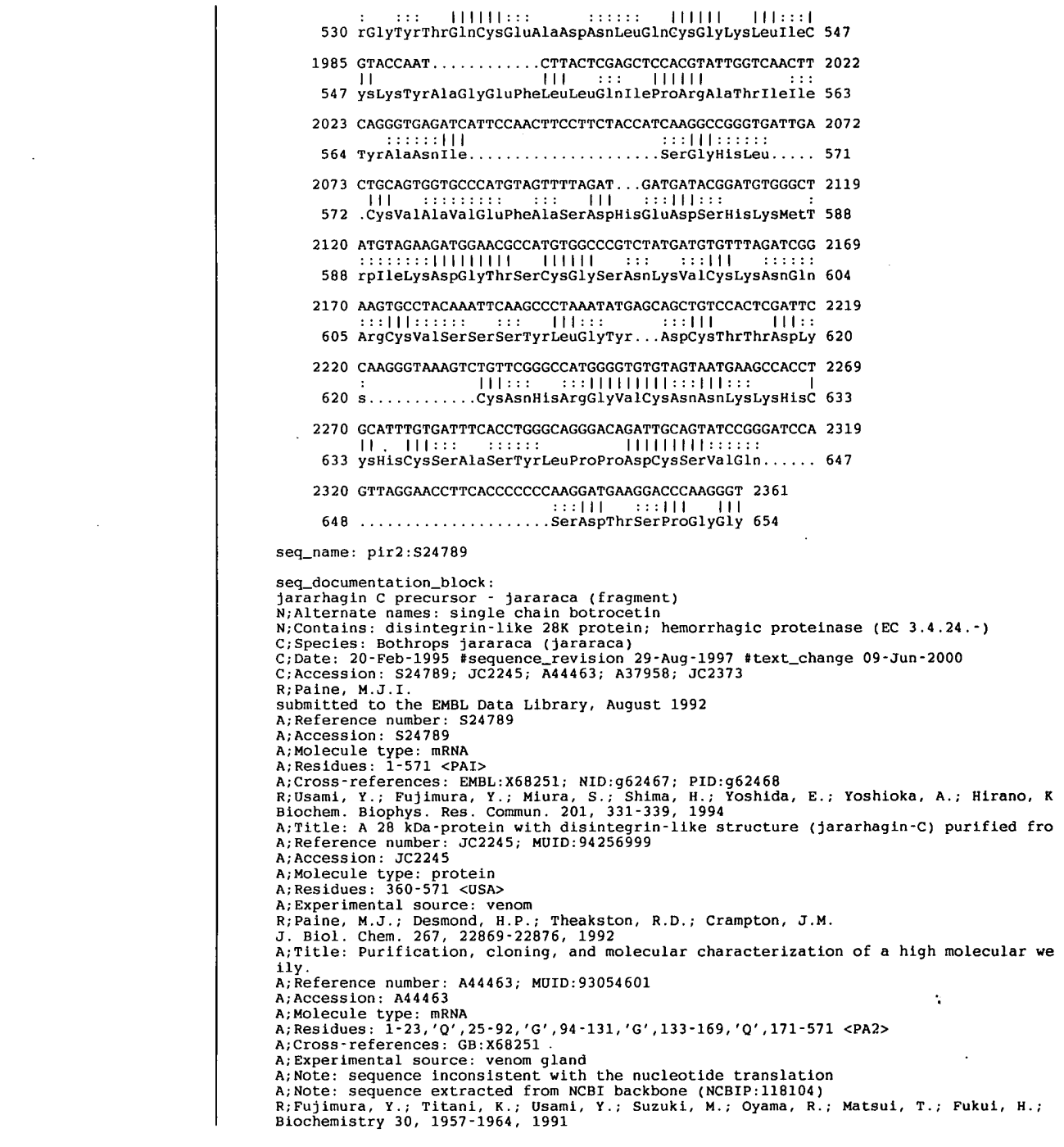
 901 TTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTC 950
     :::|||:::   :::|||   :::   :::   |||:::         ||
 180 IleGluMetHisValValValGluLysGlnLeuTyrAsnHisMetGlySe 196

 951 TTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGG 1000
     |:::   :::   :::::::::         ::::::::::::|||::::
 196 rGlyThrThrValValThrGlnLysIlePheGlnLeuIleGlyLeuThrA 213

1001 ATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTA 1050
     :::::|||:::      :::||||||      |||:::|||   ::::::
 213 snAlaIlePhe...ValSerLeuAsnIleThrValIleLeuSerSerLeu 228

1051 GAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCA 1100
     |||   |||   :::::::::::|||      |||   :::      ::
 229 GluLeuTrpIleAspGluAsnLysIleAlaThrThrGlyAspAlaLysGl 245
```





```
seq_name: pir2:S24789

seq_documentation_block:
jararhagin C precursor - jararaca (fragment)
N;Alternate names: single chain botrocetin
N;Contains: disintegrin-like 28K protein; hemorrhagic proteinase (EC 3.4.24.-)
C;Species: Bothrops jararaca (jararaca)
C;Date: 20-Feb-1995 #sequence_revision 29-Aug-1997 #text_change 09-Jun-2000
C;Accession: S24789; JC2245; A44463; A37958; JC2373
R;Paine, M.J.I.
submitted to the EMBL Data Library, August 1992
A;Reference number: S24789
A;Accession: S24789
A;Molecule type: mRNA
A;Residues: 1-571 <PAI>
A;Cross-references: EMBL:X68251; NID:g62467; PID:g62468
R;Usami, Y.; Fujimura, Y.; Miura, S.; Shima, H.; Yoshida, E.; Yoshioka, A.; Hirano, K
Biochem. Biophys. Res. Commun. 201, 331-339, 1994
A;Title: A 28 kDa-protein with disintegrin-like structure (jararhagin-C) purified fro
A;Reference number: JC2245; MUID:94256999
A;Accession: JC2245
A;Molecule type: protein
A;Residues: 360-571 <USA>
A;Experimental source: venom
R;Paine, M.J.; Desmond, H.P.; Theakston, R.D.; Crampton, J.M.
J. Biol. Chem. 267, 22869-22876, 1992
A;Title: Purification, cloning, and molecular characterization of a high molecular we
ily.
A;Reference number: A44463; MUID:93054601
A;Accession: A44463
A;Molecule type: mRNA
A;Residues: 1-23,'Q',25-92,'G',94-131,'G',133-169,'Q',171-571 <PA2>
A;Cross-references: GB:X68251
A;Experimental source: venom gland
A;Note: sequence inconsistent with the nucleotide translation
A;Note: sequence extracted from NCBI backbone (NCBIP:118104)
R;Fujimura, Y.; Titani, K.; Usami, Y.; Suzuki, M.; Oyama, R.; Matsui, T.; Fukui, H.;
Biochemistry 30, 1957-1964, 1991
```

```
A;Title: Isolation and chemical characterization of two structurally and functionally di
A;Reference number: A37958; MUID:91129280
A;Accession: A37958
A;Molecule type: protein
A;Residues: 360-372,'E',374-378,'X',380-384 <FUJ>
A;Note: 361-Val was also found
C;Comment: Inhibits collagen- and ADP-induced platelet aggregation.
C;Superfamily: mouse meltrin alpha; disintegrin homology
C;Keywords: hydrolase; metalloproteinase; venom; zinc
F;360-571/Product: jararhagin C #status experimental <MAT>
F;362-444/Domain: disintegrin homology <DIS>
F;295,299,305/Binding site: zinc (His) #status predicted
F;296/Active site: Glu #status predicted


alignment_scores:
            Quality:  751.50              Length:    617
              Ratio:  2.020                Gaps:     23
 Percent Similarity:  60.292   Percent Identity: 30.146

[redacted]ment_block:
[redacted]9-634-252A-2 x S24789   ..

Align seg 1/1  to: S24789  from: 1  to: 571

    403 AAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGC 452
        |||:::      |||   |||::::::   |||:::
      5 LysGlyAlaValGlnProLysTyrGluAspAlaMet..............  16

    453 AAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATAC 502
         ::::::::::::::::::   |||:::      :::|||   |||
     17 .GlnTyrGluPheLysValAsnGlyGluProValValLeuHisLeuGluL 33

    503 TGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAA... 549
          |||:::|||||||:::|||:::|||||||   ||||||||||||:::
     33 ysAsnLysGlyLeuPheSerLysAspTyrSerGluIleHisTyrSerPro 49

    550 AATGGGAAA......CCACAGTACTCTAAGGGTGGAGAGCACTGTTACTA 593
        :::|||:::            |||            :::||||||||||
     50 AspGlyArgGluIleThrThrTyrProProValGluAspHisCysTyrTy 66

    594 CCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCT 643
        |||||||   |||::::::      ||||||   ::::::::::|||:::|
     66 rHisGlyArgIleGluAsnAspAlaAspSerThrAlaSerIleSerAlaC 83

    644 GCAATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATA 693
        ||||||||||   |||   |||:::   :::      :::|||:::|||
     83 ysAsnGlyLeuLysGlyTyrPheLysLeuGlnArgGluThrTyrPheIle 99

    694 GAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCACATAT 743
        |||||||||:::|||      :::|||                  |||
    100 GluProLeuLysLeuProAspSerGlu.............AlaHisAl 111

    744 AATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCA 793
         :::                                 :::|||      :
    111 aVal.............................PheLysTyrGluA 117

    794 CTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTA......CAG 837
        :::::|||:::   |||:::  |||   :::      :::      :::
    117 snValGluLysGluAspGluAlaProLysMetCysGlyValThrGlnAsn 133

    838 TGG.................TTGAAAAGAAGGAAGAGAGCAGTGAATCC 869
        |||                    |||      :::   |||   :::
    134 TrpLysSerTyrGluProIleLysLysAlaSerGlnLeuAlaPheThrAl 150

    870 ATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTA 919
         ::::::     ::::::      |||||||:::|||:::::::::|||
    150 aGluGlnGlnArgTyrAspProTyrLysTyrIleGluPhePheValValV 167

    920 ATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAAC 969
          |||:::   |||      |||:::::::::::            :::
    167 alAspGlnGlyThrValThrLysAsnAsnGlyAspLeuAspLysIleLys 183

    970 AACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCA 1019
                    :::::::::|||:::|||::::::|||::::::
    184 AlaArgMetTyrGluLeuAlaAsnIleValAsnGluIlePheArg...Ty 199

   1020 GCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGG 1069
         |||      :::|||:::||||||:::::|||   |||::::::   |
    199 rLeuTyrMetHisValAlaLeuValGlyLeuGluIleTrpSerAsnGlyA 216

   1070 ATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCA 1119
        ||:::|||   :::      :::            |||:::::::|||:::
    216 spLysIleThrValLysProAspValAspTyrThrLeuAsnSerPheAla 232

   1120 AAATACCGGCAG.............CGCATTAAGCAGCATGCTGATGCTGT 1157
        ::::::|||:::                  |||:::|||   |||   ::
    233 GluTrpArgLysThrAspLeuLeuThrArgLysLysHis...AspAsnAl 248

   1158 GCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTT 1207
        ::::|||::::::   :::   |||:::            :::   :::|
    248 aGlnLeuLeuThrAlaIleAspPheAsnGlyProThrIleGlyTyrAlaT 265

   1208 ACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGT...GTGAATGAG 1254
        ||:::|||::::::|||         |||:::||||||   |||::::::
    265 yrIleGlySerMetCysHisProLysArgSerValGlyIleValGlnAsp 281

   1255 TATGGT...CTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGC 1301
        |||:::   :::   ::::::||||||   ::::::::::::::::::::
    282 TyrSerProIleAsnLeuValValAlaValIleMetAlaHisGluMetGl 298

   1302 TCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATGTG 1351
        ::::|||||||||||:::   :::   :::|||             |||:
    298 yHisAsnLeuGlyIleHisHisAspThrGlySer.............CysS 311

   1352 ACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCC 1401
        ::|||   :::              |||||||||      |||   :::|||
    311 erCysGlyAspTyr......ProCysIleMetGlyProThr...IleSer 324

   1402 CAT......TCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGA 1445
        :::      |||:::   ||||||:::||||||   ::::::      ||
    325 AsnGluProSerLysPhePheSerAsnCysSerTyrIleGlnCysTrpAs 341

   1446 CTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCA......A 1489
        ||||:::::::::   :::      |||:::::::|||:::|||      |
    341 pPheIleMetAsnHisAsnProGluCysIleIleAsnGluProLeuGlyT 358

   1490 CAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGG 1539
        ||   :::::::::|||      |||||||||      :::|||:::|||
    358 hrAspIleIleSerProProValCysGlyAsnGluLeuLeuGluValGly 374

   1540 GAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTATGGATTATGCTGT.. 1587
        ||||||||||||||||||         :::|||   :::   ||||||
    375 GluGluCysAspCysGlyThrProGluAsnCysGlnAsnGluCysCysAs 391

   1588 ....AAGAAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCT 1633
                |||:::|||:::::::|||:::::::|||:::   |||   |
    391 pAlaAlaThrCysLysLeuLysSerGlySerGlnCysGlyHisGlyAspC 408

   1634 GCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGG 1683
        |||||:::      :::|||   |||:::      |||   |||||||||
    408 ysCysGlu......GlnCysLysPheSerLysSerGlyThrGluCysArg 422

   1684 GATGCTGTGAACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGG 1733
          :::::::::|||||||||   :::|||:::|||||||||:::|||::
    423 AlaSerMetSerGluCysAspProAlaGluHisCysThrGlyGlnSerSe 439

   1734 TCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAA 1783
        ::::||||||   ::::::      ::::::|||      |||   :::|
    439 rGluCysProAlaAspValPheHisLysAsnGlyGlnProCysLeuAspA 456

   1784 ATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGT 1833
        ||   |||   ||||||||||||:::|||            :::||||||
    456 snTyrGlyTyrCysTyrAsnGlyAsnCysProIleMetTyrHisGlnCys 472

   1834 CAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGA 1883
```

```
          :::::::|||:::   :::      ::::::      |||:::
    473 TyrAlaLeuPheGlyAlaAspValTyrGluAlaGluAspSerCysPhe.. 488

   1884 AAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGCAG 1933
         |||   |||   :::|||:::      |||   |||   |||::::::
    489 .LysAspAsnGlnLysGlyAsnTyrTyrGlyTyrCysArgLysGluAsnG 505

   1934 ACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTC 1983
        :::   |||   |||:::   :::||||||   ||||||   |||
    505 lyLysLysIleProCysAlaProGluAspValLysCysGlyArgLeuTyr 521

   1984 TGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGAT 2033
        |||   :::      ::::::|||                  |||:::
    522 CysLysAsp......AsnSerPro.................GlyGlnAs 530

   2034 CATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTG 2083
           |||   ::::::::::::::::
    530 nAsnProCysLysMetPheTyrSer..................... 538

   2084 CCCATGTAGTTTTAGATGATGATACGGATGTGGGCTATGTAGAAGATGGA 2133
                     :::||||||         |||   |||      |||
    539 .............AsnAspAspGluHisLysGlyMetValLeuProGly 550

   2134 ACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAAT 2183
        |||   |||:::   :::   :::|||   :::      |||::::::::
    551 ThrLysCysAlaAspGlyLysValCysSerAsnGlyHisCysValAspVa 567

   2184 T 2184
        :
    567 l 567

seq_name: pir2:T26644

seq_documentation_block:
hypothetical protein Y37D8A.13 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T26644
R;Barlow, K.
submitted to the EMBL Data Library, October 1998
A;Reference number: Z20247
A;Accession: T26644
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1042 <WIL>
A;Cross-references: EMBL:AL032626; PIDN:CAA21545.1; CESP:Y37D8A.13
A;Experimental source: clone Y37D8A
A;Genetics:
A;Gene: CESP:Y37D8A.13
A;Introns: 25/1; 239/3; 263/3; 433/3; 859/2


alignment_scores:
            Quality:  746.50              Length:    740
              Ratio:  1.752                 Gaps:     26
 Percent Similarity:  57.568   Percent Identity: 27.838

alignment_block:
 US-09-634-252A-2 x T26644   ..

 Align seg 1/1  to: T26644  from: 1  to: 1042

    382 CCTTATCACGTTCTTGACACAAAGGCAAGA..................... 411
        |||:::::::::   |||   :::   |||
     50 ProPheGlnIleArgAspLysAsnGluArgIleGlyIleAspThrArgAs 66

    412 ..........CACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGG 451
                      |||:::|||   :::   |||   ::::::
     66 nTyrPheLeuLysAlaGlnGluHisTyrSerHisValThrIleValIleA 83

    452 CAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCAAA 501
          |||   |||:::      :::      |||::::::|||:::
     83 rgSerAsnGlnLeuGlyArgLeu......LysLeuValLeuGlu..... 95

    502 CTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTAC..... 546
           :::|||:::::::::      :::   :::   :::|||
     96 ...ArgAsnAsnPheIlePheLeuAsnGlnThrAlaPheHisLysLeuAs 111

    547 .GAAAATGGGAAACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACC 595
           :::|||:::         ::::::      |||:::|||||||||:.
    111 pAlaAspGlyGluArgValIleGlnAsnArgValGluAsnCysTyrTyrG 128

    596 ATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGC 645
        ::|||::::::   |||   ::::::|||   ||||||||||||:::|||
    128 lnGlyThrValGlyGlyGluGluSerSerPheValAlaLeuSerSerCys 144

    646 AATGGACTTCATGGCATG......TTTGAAGATGATACCTTCGTGTATAT 689
        |||||||||:::|||:::      |||   :::   |||   ::::::
    145 AsnGlyLeuArgGlyValIleSerPheAlaAsnGlyThr...ThrPheGl 160

    690 GATAGAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCAC 739
         |||   ||||||:::         |||:::::|||      :::||||
    160 yIleTrpProLeuAsp...GlyGlyAspArgAsnSerArgArgHisProH 176

    740 ATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAAT 789
        |||||:::   |||:::      :::|||   :::|||      ::::::
    176 isIleLeuTyrLysSerGluTrpSerGlnGluAlaLysCysGlySerSer 192

    790 CTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTG 839
        ::::::         |||:::
    193 MetAlaHisAlaValGlyGln.......................... 199

    840 GTTGAAAAGAAGGAAGAGAGCAGTG..................... 864
           :::|||   |||:::   |||
    200 ....ArgArgMetLysLysHisValHisLysHisArgSerHisHisHisG 215

    865 .....AATCCATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTT 909
           |||   :::|||   :::   ::::::   ||||||:::|||:::
    215 luHisAsnLysLysArgAspValSerLysArgThrLysTyrValGluVal 231

    910 ATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGC 959
        ::::::   |||::::::      |||   :::
    232 AlaLeuIleAlaAspTyrGluPheMetLysAlaArgGlyLeuHisAspLe 248

    960 ACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTT 1009
           :::   ::::::      :::   :::|||::::::||||||:::
    248 uAspAlaIleSerTyrMetLeuGluSerLeuAsnIleAlaAspSerMetL 265

   1010 ACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGG 1059
           :::::::::||||||   |||:::      |||   ||||||   |||
    265 euSerArgAspLeuAsnIleArgLeuSerAlaValTyrValGluLeuTrp 281

   1060 ACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCA 1109
        |||:::   ::::::||||||:::      :::      :::   |||
    282 ThrAspValGlnArgIleAspLeuTrpGluAspIleGluArgThrLeuSe 298

   1110 TGAGTTCTCAAAATACCGGCAGCGCATTAAGCAGCAT......GCTGATG 1153
                   |||   :::         |||         ||||
    298 rGlyValValAspTyrAlaSerGlyHisIleTyrHisIleGlnLysAspA 315

   1154 CTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTG 1203
        ||      |||::::::      :::|||      ::::::   |||
    315 laSerIleLeuPheThrAlaGlySerPheAlaAsnGlnGluValSerAsn 331

   1204 AGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTT.........GG 1244
        :::   :::   ::::::|||:::          :::|||          ||
    332 AlaAlaIleArgSerIleCysThrAlaArgSerAlaValIleValLysGl 348

   1245 TGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAGA 1294
        ||||::::::::::::                 :::::::::|||:::||||
    348 yValGluGlnPheAlaThrHisTrpAsn...GlyGluLeuLeuAlaGlnS 364

   1295 GCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCA 1344
        ||:::::::::   ||||||::::::   :::   ::::::
    364 erIleGlyHisLeuLeuGlyLeuGluHisAspThrThrAla......... 377
```

```
1345 AAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACA.. 1392
      |||:::|||      |||         |||:::|||:::::
 378 ...CysSerCysGluProSer...ProGluCysValMetArgGlnGlnPr 392

1393 .....GGGGTGTCCCATTCTCGAAAATTTTCAA 1420
                        |||         |||   :::|||||||
 392 oGlyArgValGlyGlyGlyGlySerProPheSerTrpGlnPheSerL 409

1421 AGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCC 1470
     ||||||||:::   :::   :::  :::   |||   |||:::
 409 ysCysSerValAlaArgMetHisGlyIleTrpGlnAspGlyAsnIleGln 425

1471 TGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAA 1520
     ||||||:::|||:::|||   ::::::   |||      ||||||||||
 426 CysLeuLeuAsnLysProPheGlnValSerGluLeuArgGluCysGlyAs 442

1521 TGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAAT 1570
     ||||   |||:::::::::|||||||||||||||||||   :::   :::|
 442 nGlyValValAspGlySerGluGluCysAspCysGlySerArgGluAsnC 459

1571 GCTATGGATTATGCTGTAAG......AAATGTTCCCTCTCCAACGGGGCT 1614
     ||        ||||||          |||:::|||         |||
 459 ysGlnAspProCysCysAspProLeuThrCysThrLeuArgProHisAla 475

1615 CACTGC...AGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCA 1661
     :::|||   :::         ||||||:::::      |||   :::::
 476 GlnCysAlaAlaHisHisLysCysCysHisArg......CysGluLeuAr 490

1662 GCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTG 1711
     :      |||      ||||||:::::   :::   |||||||:::::|
 490 gLysAlaGlyAspThrCysArgSerSerLysSerProCysAspValAlaG 507

1712 AATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAA 1761
     ||   |||   |||   ||||||:::|||||||||:::   |||
 507 luGlnCysAspGlyLysSerGlyAspCysProProAspGlyHisLeuIle 523

1762 GACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTG 1811
     ||||||   :::|||:::   :::::::|||:::|||:::::::|||:::||
 524 AspGlyThrValCysGly...ThrAspGlyGlnCysTrpArgGlyAsnCy 539

1812 CAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAG 1861
     |:::         :::|||||||||   :::||||||   :::|||
 539 sSerAspSerHisGlnGlnCysGlnLysLeuTrpGlyArgGluAlaArgV 556

1862 GGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAG 1911
       ::::::      |||:::|||:::   ||||||:::|||:::|||
 556 alAlaGluProValCysPheGluGln...AsnThrLysGlyAlaGluTyr 571

1912 GGAAACTGCGGGAAG.........GATGGAGACCGGTGGATTCAGTGCAG 1952
     :::||||||||:::         ||||||      :::      |||::
 572 AlaAsnCysGlyGlnArgGlnAlaAspGly...ThrTyrHisProCysGl 587

 953 CAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAG 2002
     :   :::|||:::   ||||||   |||   |||   :::   :::   :
 587 nIleGluAspThrArgCysGlyThrLeuHisCysHisSerGlySerIleT 604

2003 CTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTAC 2052
     ::|||      ::::::|||::::::       :::         |||
 604 hrProIleAspSerSerLeuLysAla......PheThrPheHisPheThr 618

2053 CATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGA 2102
     ::::::::::::   |||:::|||::::::
 619 GluAsnSerHisGlnIleGlnCysLysSerIle..............Al 630

2103 TGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTA 2152
      ::::::   ||||||   ::::::|||||||||   |||:::   :::
 630 aSerAlaAlaValGlyLeuThrSerAspGlyThrAsnCysAlaSerGlyA 647

2153 TGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTA...AATATG 2199
       :::|||:::      :::|||:::::::::::::::::::   :::
 647 rgValCysValAlaGlySerCysValGluMetSerSerValSerSerAla 663

2200 AGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGT 2249
```

```
     ::::::||||||   ::::::   :::   :::|||||||||||||||
 664 ThrAlaCysProThrAsnAsnLeuAlaLeuLeuCysSerGlyHisGlyHi 680

2250 GTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAG 2299
      |||::::::   |||   |||:::|||         |||:::|||:::
 680 sCysThrThrThrAlaArgCysValCysPheAsnGlyTrpSerGlyValA 697

2300 ATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAAGGATGAA 2349
       |||:::|||||||:::         :::   :::
 697 laCysAspIleArgSerAsnSerSerThrTyrGlnGlySerMetGlyPhe 713

2350 GGACCCAAGGGTCCTAGTGCCACC..... 2373
     |||   :::|||   :::::::::
 714 GlyGluGluGlySerGlyGlySerSerGlnLysSerSerGluArgLysTh 730

2374 .....AATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGG 2410
                  :::|||   ||||||:::         :::      :
 730 rIleMetIleProHisLeuAsnIleGlyThrThrLeuGluThrAlaThrL 747

2411 TAGCAGCTATTGTCCTTGGG 2430
     ::   ||||||:::||||||
 747 euPheAlaIleLeuLeuGly 753
```

seq_name: pir2:JC4861

seq_documentation_block:
fertillin beta chain - human
C;Species: Homo sapiens (man)
C;Date: 15-Aug-1996 #sequence_revision 18-Oct-1996 #text_change 21-Jan-2000
C;Accession: JC4861
R;Gupta, S.K.; Alves, K.; O'Neil Palladino, L.; Mark, G.E.; Hollis, G.F.
Biochem. Biophys. Res. Commun. 224, 318-326, 1996
A;Title: Molecular cloning of the human fertilin beta subunit.
A;Reference number: JC4861; MUID:96295488
A;Accession: JC4861
A;Molecule type: mRNA
A;Residues: 1-734 <GUP>
A;Cross-references: GB:U38805; NID:g4151118; PIDN:AAD04206.1; PID:g4151119
C;Comment: This protein is an integral sperm membrane glycoprotein, and plays a role
C;Superfamily: mouse meltrin alpha; disintegrin homology
C;Keywords: glycoprotein; integrin binding; transmembrane protein
F;382-734/Product: fertillin beta chain #status predicted <NAT>
F;382-467/Domain: disintegrin homology <DIS>
F;448-450/Region: integrin binding #status predicted
F;686-708/Domain: transmembrane #status predicted <TMM>
F;121,219,352,458,565/Binding site: carbohydrate (Asn) (covalent) #status predicted

alignment_scores:
Quality: 739.00 Length: 695
Ratio: 1.852 Gaps: 23
Percent Similarity: 57.410 Percent Identity: 27.482

alignment_block:
US-09-634-252A-2 x JC4861 ..

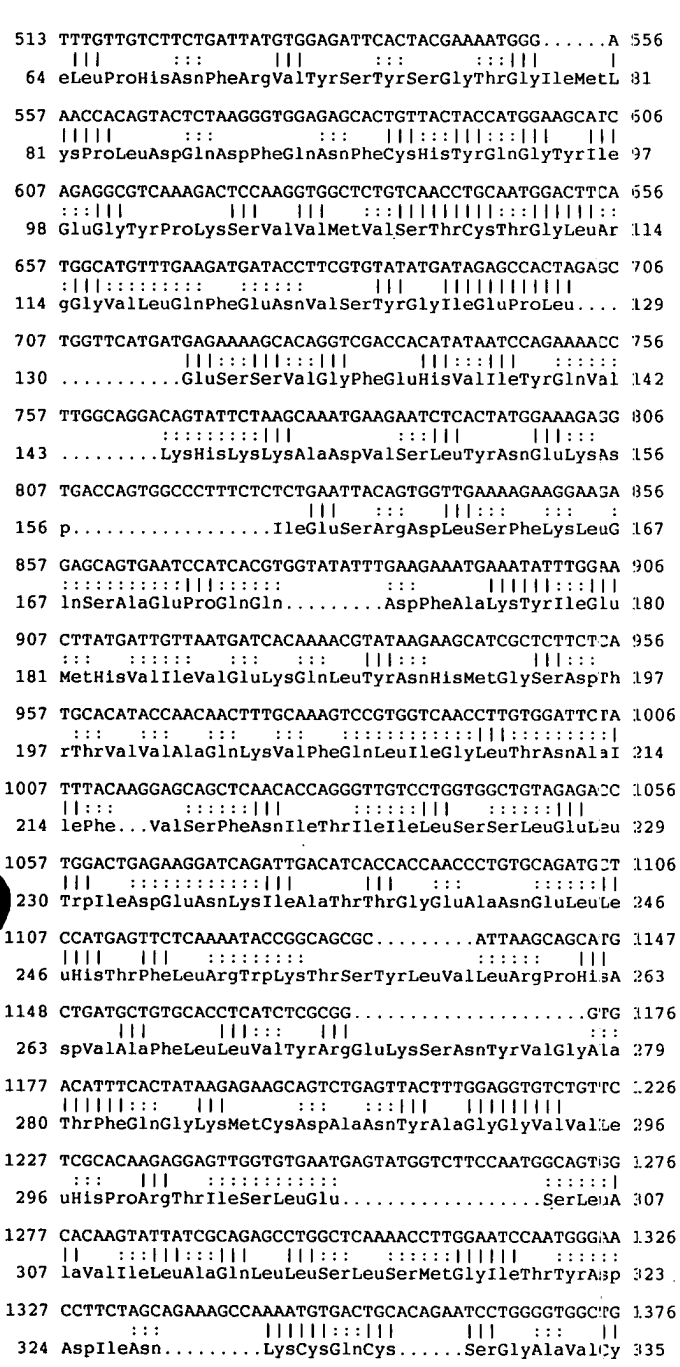
Align seg 1/1 to: JC4861 from: 1 to: 734

```
 313 AATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACAT 362
     :::::::::      :::||||||:::|||::::::::::         ||
  20 AspSerLeuProValGlnIleThrValProGluLysIleArgSerIleIl 36

 363 CAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGAC 412
     |:::::      ||||||
  36 eLysGluGlyIleGluSer..............................  42

 413 ACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAG 462
                                         |||||||||:::::
  43 .....................................GlnAlaSerTyrLys  47

 463 ATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGG 512
     |||       |||:::   :::   ::::::|||:::   :::|||
  48 IleValIleGluGlyLysProTyrThrValAsnLeuMetGlnLysAsnPh 64
```





```
          |||:::   :::|||||||||:::|||:::      |||   |||:
    620 ....CysAsnAspArgGlyValCysAsnAsnLysLysHisCysHisCysS 635

   2279 ATTTCACCTGGGCAGGGACAGATTGCAGTATCCGG 2313
        ::   ::::::         |||||||||::::::
    635 erAlaSerTyrLeuProProAspCysSerValGln 646

seq_name: pir2:A55796

seq_documentation_block:
ecarin precursor - saw-scaled viper
C;Species: Echis carinatus (saw-scaled viper)
C;Date: 05-Jan-1996 #sequence_revision 05-Jan-1996 #text_change 21-Jul-2000
C;Accession: A55796
R;Nishida, S.; Fujita, T.; Kohno, N.; Atoda, H.; Morita, T.; Takeya, H.; Kido, I.; Paine
Biochemistry 34, 1771-1778, 1995
A;Title: cDNA cloning and deduced amino acid sequence of prothrombin activator (ecarin)
A;Reference number: A55796; MUID:95151760
A;Accession: A55796
A;Status: preliminary
[redacted]ecule type: mRNA
[redacted]idues: 1-616 <NIS>
[redacted]oss-references: GB:D32212; NID:g717090; PIDN:BAA06910.1; PID:g717091
C;Superfamily: mouse meltrin alpha; disintegrin homology
F;404-486/Domain: disintegrin homology <DIS>
F;338/Active site: Glu #status predicted


alignment_scores:
            Quality:  735.00              Length:    660
              Ratio:  1.885                 Gaps:     24
 Percent Similarity:  59.091   Percent Identity: 28.182

alignment_block:
 US-09-634-252A-2 x A55796   ..

 Align seg 1/1  to: A55796  from: 1  to: 616

    283 CAGAATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAAT 332
        |||:::   |||      ::::::   :::|||:::   :::   |||::
     17 GlnGlyCysSerIleIleLeuGlySerGlyAsnValAsnAspTyrGluVa 33

    333 CACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGAAAGCC 382
        ::::   |||:::::::::          :::          ::::::|
     33 lValTyrProGlnLysValThrAlaLeuProLysGlyAlaValGlnGlnP 50

    383 CTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAG 432
        ||                           :::||||||::::::
     50 ro..........................GluGlnLysTyrGluAsp 56

    433 GCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAA 482
        |||:::                 ::::::::::::|||:::   |||:::
     57 AlaMet..............GlnTyrGluPheGluValLysGlyGluPr 68

    483 ATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATG 532
          :::|||   |||      |||:::   |||:::|||:::||||||
     68 oValValLeuHisLeuGluLysAsnLysGluLeuPheSerGluAspTyrS 85

    533 TGGAGATTCACTACGAAAATGGGAAACCACAGTACTCTAAGGGT...... 576
          |||   ||||||::::::          :::   :::   :::
     85 erGluThrHisTyrSerSerAspAspArgGluIleThrThrAsnProSer 101

    577 ...GGAGAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTC 623
           :::||||||||||||||||   |||::::::      :::||
    102 ValGluAspHisCysTyrTyrHisGlyArgIleGlnAsnAspAlaGluSe 118

    624 CAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATG 673
        |   ::::::::::|||:::|||||||||||   |||   |||:::
    118 rThrAlaSerIleSerAlaCysAsnGlyLeuLysGlyHisPheLysLeuA 135

    674 ATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAA 723
             :::|||:::||||||||||||:::::::      :::|||
    135 rgGlyGluThrTyrPheIleGluProLeuLysIleProAspSerGlu... 150

    724 AGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCAGGACAGTATTC 773
                      |||   :::   |||                |||::
    151 ............AlaHisAlaValTyrLys..............TyrGl 158

    774 TAAGCAAATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTC 823
        :                  ::::::|||:::   |||:::   |||   :
    158 u..................AsnIleGluAsnGluAspGluAlaProLysM 169

    824 TC.........TCTGAATTACAGTGG..............TTGAAAAGA 849
        ::         ::::::   :::|||                :::|||:::
    169 etCysGlyValThrGlnAspAsnTrpGluSerAspGluProIleLysLys 185

    850 AGGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATA 899
                  :::   |||   :::             |||   |||::
    186 ThrLeuGlyLeuIleValProProHisGluArgLysPheGluLysLysPh 202

    900 TTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCT 949
        ::::||||||::::::|||   ||||||:::          |||:::::::
    202 eIleGluLeuValValValValAspHisSerMetValThrLysTyrAsnA 219

    950 CTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTG 999
        :::::   :::      :::::::::      :::::::::|||   |||
    219 snAspSerThrAlaIleArgThrTrpIleTyrGluMetLeuAsnThrVal 235

   1000 GATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGT 1049
        ::::::||||||         :::|||   ||||||:::||||||:::::
    236 AsnGluIleTyrLeu...ProPheAsnIleArgValAlaLeuValGlyLe 251

   1050 AGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGC 1099
        :|||   |||   :::   |||   |||:::::::|||::::::      :
    251 uGluPheTrpCysAsnGlyAspLeuIleAsnValThrSerThrAlaAspA 268

   1100 AGATGCTCCATGAGTTCTCAAAATACCGGCAG.............CGCATT 1137
        ::   ||||||:::|||:::::::::|||                :::
    268 spThrLeuHisSerPheGlyGluTrpArgAlaSerAspLeuLeuAsnArg 284

   1138 AAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTA 1187
        |||:::|||   |||   :::::::|||:::::::::|||||||:::   ::
    285 LysArgHis...AspHisAlaGlnLeuLeuThrAsnValThrLeuAspHi 300

   1188 TAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAG 1237
        ::::      ::::::::::::      |||:::|||:::      |||:
    300 sSerThrLeuGlyIleThrPheValTyrGlyMetCysLysSerAspArgS 317

   1238 GAGTTGGT...GTGAATGAGTAT...GGTCTTCCAATGGCAGTGGCACAA 1281
        ::|||      :::   :::||   ::::::   :::   :::|||
    317 erValGluLeuIleLeuAspTyrSerAsnIleThrPheAsnMetAlaTyr 333

   1282 GTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTC 1331
        ::::::::::::::::::::::::::::::||||||:::      :::   ::
    334 IleIleAlaHisGluMetGlyHisSerLeuGlyMetLeuHisAspThrLy 350

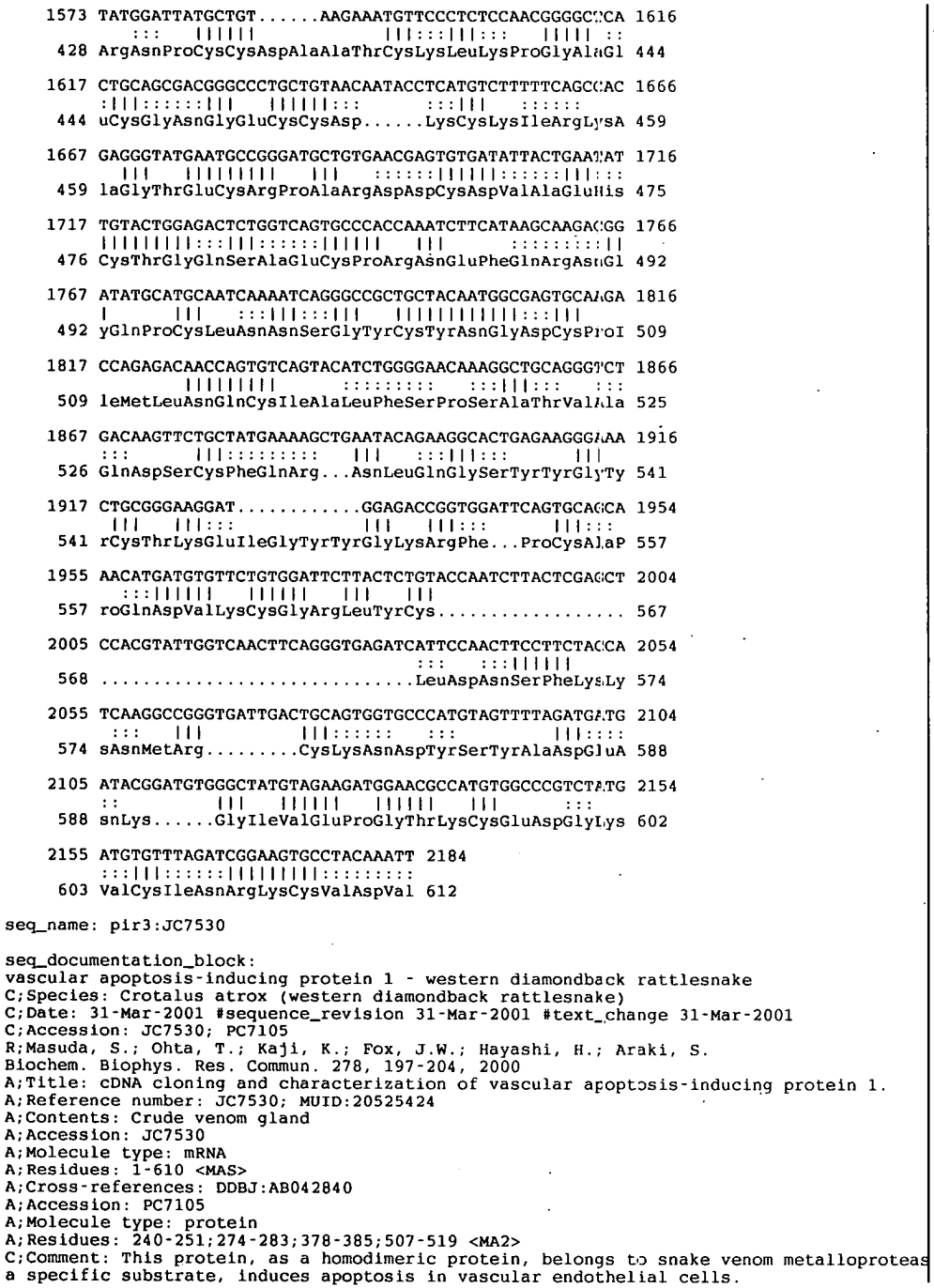
   1332 TAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCA 1381
        :                     ||||||   :::       |||||||
    350 sPhe..................CysThrCysGlyAlaLysProCysIleM 361

   1382 TG...GAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGC 1428
        ||      :::   ::::::         ::::::|||||:::|||||||
    361 etPheGlyLysGluSerIleProProProLysGluPheSerSerCysSer 377

   1429 ATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTT 1478
              :::|||:::   :::|||   :::   :::       |||:::::
    378 TyrAspGlnTyrAsnLysTyrLeuLeuLysTyrAsnProLysCysIleLe 394

   1479 CAACAGGCCAACAAAG......CTATTTGAGCCCACGGAATGTGGAAATG 1522
        ::::   |||   :::       :::   :::|||:::   |||||||||
    394 uAspProProLeuArgLysAspIleAlaSerProAlaValCysGlyAsnG 411

   1523 GATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGC 1572
              |||   |||||||||||||||||||      :::::|||
    411 luIleTrpGluGluGlyGluGluCysAspCysGlySerProAlaAspCys 427
```



```
 1573 TATGGATTATGCTGT......AAGAAATGTTCCCTCTCCAACGGGGCCCA 1616
 428 ArgAsnProCysCysAspAlaAlaThrCysLysLeuLysProGlyAlaGl 444

 1617 CTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCAC 1666
 444 uCysGlyAsnGlyGluCysCysAsp......LysCysLysIleArgLysA 459

 1667 GAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATAT 1716
 459 laGlyThrGluCysArgProAlaArgAspAspCysAspValAlaGluHis 475

 1717 TGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGG 1766
 476 CysThrGlyGlnSerAlaGluCysProArgAsnGluPheGlnArgAsnGl 492

 1767 ATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGA 1816
 492 yGlnProCysLeuAsnAsnSerGlyTyrCysTyrAsnGlyAspCysProI 509

 1817 CCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCT 1866
 509 leMetLeuAsnGlnCysIleAlaLeuPheSerProSerAlaThrValAla 525

 1867 GACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAA 1916
 526 GlnAspSerCysPheGlnArg...AsnLeuGlnGlySerTyrTyrGlyTy 541

 1917 CTGCGGGAAGGAT............GGAGACCGGTGGATTCAGTGCAGCA 1954
 541 rCysThrLysGluIleGlyTyrTyrGlyLysArgPhe...ProCysAlaP 557

 1955 AACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCT 2004
 557 roGlnAspValLysCysGlyArgLeuTyrCys..................  567

 2005 CCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCA 2054
 568 ............................LeuAspAsnSerPheLysLy 574

 2055 TCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATG 2104
 574 sAsnMetArg.........CysLysAsnAspTyrSerTyrAlaAspGluA 588

 2105 ATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATG 2154
 588 snLys......GlyIleValGluProGlyThrLysCysGluAspGlyLys 602

 2155 ATGTGTTTAGATCGGAAGTGCCTACAAATT 2184
 603 ValCysIleAsnArgLysCysValAspVal 612
```

seq_name: pir3:JC7530

seq_documentation_block:
vascular apoptosis-inducing protein 1 - western diamondback rattlesnake
C;Species: Crotalus atrox (western diamondback rattlesnake)
C;Date: 31-Mar-2001 #sequence_revision 31-Mar-2001 #text_change 31-Mar-2001
C;Accession: JC7530; PC7105
R;Masuda, S.; Ohta, T.; Kaji, K.; Fox, J.W.; Hayashi, H.; Araki, S.
Biochem. Biophys. Res. Commun. 278, 197-204, 2000
A;Title: cDNA cloning and characterization of vascular apoptosis-inducing protein 1.
A;Reference number: JC7530; MUID:20525424
A;Contents: Crude venom gland
A;Accession: JC7530
A;Molecule type: mRNA
A;Residues: 1-610 <MAS>
A;Cross-references: DDBJ:AB042840
A;Accession: PC7105
A;Molecule type: protein
A;Residues: 240-251;274-283;378-385;507-519 <MA2>
C;Comment: This protein, as a homodimeric protein, belongs to snake venom metalloproteas
a specific substrate, induces apoptosis in vascular endothelial cells.

C;Genetics:
A;Gene: vapl
C;Keywords: apoptosis

alignment_scores:
Quality: 725.00 Length: 657
Ratio: 1.893 Gaps: 24
Percent Similarity: 58.295 Percent Identity: 28.919

alignment_block:
US-09-634-252A-2 x JC7530 ..

Align seg 1/1 to: JC7530 from: 1 to: 610

```
  283 CAGAATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAAT 332
   17 GlnGlySerSerValIleLeuGluSerGlyAsnValAsnAspTyrGluVa 33

  333 CACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGAAAGCC 382
   33 lValTyrProArgLysValThrAlaLeu..................... 42

  383 CTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAG 432
   43 .................ProLysGlyAlaValGlnProLysTyrGluAsp 53

  433 GCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAA 482
   54 AlaMet..............GlnTyrGluPheLysValAsnGlyGluPr 65

  483 ATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATG 532
   65 oValValLeuHisLeuGluLysAsnLysGlyLeuPheSerGluAspTyrS 82

  533 TGGAGATTCACTACGAA...AATGGGAAA......CCACAGTACTCTAAG 573
   82 erGluThrHisTyrSerProAspGlyArgGluIleThrThrTyrProPro 98

  574 GGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTC 623
   99 ValGluAspHisCysTyrTyrHisGlyArgIleGluAsnAspAlaAspSe 115

  624 CAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATG 673
  115 rThrAlaSerIleSerAlaCysAsnGlyLeuLysGlyHisPheLysLeuG 132

  674 ATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAA 723
  132 lnGlyGluMetTyrLeuIleGluProLeuLysLeuProAspSerGlu... 147

  724 AGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCAGGACAGTATTC 773
  148 ............AlaHisAlaVal........................... 151

  774 TAAGCAAATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTC 823
  152 .......PheLysTyrGluAsnValGluLysGluAspGluAlaProLysM 166

  824 TCTCTGAATTA......CAGTGG..............TTGAAAAGAAGG 852
  166 etCysGlyValThrGlnAsnTrpGluSerTyrGluProIleLysLysAla 182

  853 AAGAGAGCA...GTGAATCCATCACGTGGTATATTT...GAAGAAATGAA 896
  183 SerGlnSerAsnLeuThrProGluGlnGlnArgTyrLeuAsnAlaLysLy 199

  897 ATATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATC 946
  199 sTyrValLysLeuPheLeuValAlaAspTyrIleMetTyrLeuLysTyrG 216

  947 GCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTT 996
```

```
 216 lyArgAsnLeuThrAlaValArgThrArgMetTyrAspIleValAsnVal 232

 997 GTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGC 1046
     :::      ||||||   :::::::::|||   :::|||:::||||||::
 233 IleThrProIleTyr...HisArgMetAsnIleHisValAlaLeuValGl 248

1047 TGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTG 1096
     ::::|||   |||::::::   |||:::|||   :::   ::::::
 248 yLeuGluIleTrpSerAsnThrAspLysIleIleValGlnSerSerAlaA 265

1097 TGCAGATGCTCCATGAGTTCTCAAAATACCGGCAG.....CGC 1134
            |||      |||:::|||:::|||
 265 spValThrLeuAspLeuPheAlaLysTrpArgAlaThrAspLeuLeuSer 281

1135 ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCA 1184
     |||:::|||   |||   ::::::|||::::::   ::::::|||::
 282 ArgLysSerHis...AspAsnAlaGlnLeuLeuThrGlyIleAsnPheAs 297

1185 CTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAA 1234
     :       :::::|||:::|||:::||||||:::|||:::
 297 nGlyProThrAlaGlyLeuGlyTyrLeuGlyGlyIleCysAsnThrMetT 314

1235 GAGGAGTTGGTGTG......AATGAGTATGGTCTTCCAATGGCAGTGGCA 1278
       ::::::|||:::      ::::::      :::         ||||||
 314 yrSerAlaGlyIleValGlnAspHisSerLysIleHisHisLeuValAla 330

1279 CAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACC 1328
     :::::::::::::::::::::::::|||||||||:::::::   :::
 331 IleAlaMetAlaHisGluMetGlyHisAsnLeuGlyMetAspHisAsp.. 346

1329 TTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCA 1378
                ||| |||||| ::: |||:
 347 ..............LysAspThrCysThrCysGlyThrArgProCysV 358

1379 TCATGGAG...GAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGC 1425
     ::|||          ::: ::: |||  ||||||  |||
 358 alMetAlaGlyAlaLeuSerCysGluAlaSerPheLeuPheSerAspCys 374

1426 AGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCT 1475
     |||    ::::::|||:::||||||  ::::::      |||::
 375 SerGlnLysAspHisArgGluPheLeuIleLysAsnMetProGlnCysIl 391

1476 TTTCAACAGGCCA......ACAAAGCTATTTGAGCCCACGGAATGTGGAA 1519
     ::::::::::|||    |||  :::  :::|||:::  |||||||
 391 eLeuLysLysProLeuLysThrAspValValSerProAlaValCysGlyA 408

1520 ATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAA 1569
     ||  :::||||||:::||||||||||||||||||||
 408 snTyrPheValGluValGlyGluGluCysAspCysGlySerProArgThr 424

1570 TGCTATGGATTATGCTGT......AAGAAATGTTCCCTCTCCAACGGGGC 1613
     |||        ||||||          |||:::|||  :::||||
 425 CysArgAspProCysCysAspAlaThrThrCysLysLeuArgGlnGlyAl 441

1614 TCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGC 1663
     |:::|||:::::|||  ||||||:::      :::|||  |||:::
 441 aGlnCysAlaGluGlyLeuCysCysAsp......GlnCysArgPheLysG 456

1664 CACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAA 1713
         |||  |||||||||  |||  :::||||||||:::::::::
 456 lyAlaGlyThrGluCysArgAlaAlaLysAspGluCysAspMetAlaAsp 472

1714 TATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGA 1763
        |||||||||   |||:::::::|||      ::::::  ::::::::
 473 ValCysThrGlyArgSerAlaGluCysThrAspArgPhe...GlnArgAs 488

1764 CGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCA 1813
     :|||     |||:::::::|||:::|||   ||||||||||||:::|||
 488 nGlyGlnProCysLysAsnAsnAsnGlyTyrCysTyrAsnGlyLysCysP 505

1814 AGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGG 1863
                :::||||||      :::::::|||      |||:::
 505 roIleMetAlaAspGlnCysIleAlaLeuPheGlyProGlyAlaThrVal 521
```

```
1864 TCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGG 1913
     |||:::      |||:::::::   :::|||   ||||||::::::   ||
 522 SerGlnAspAlaCysPheGln...PheAsnArgGluGlyAsnHisTyrGl 537

1914 AAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATG 1963
     |   |||   |||:::   :::      |||   |||:::   :::||||
 537 yTyrCysArgLysGluGlnAsnThrLysIleAlaCysGluProGlnAspV 554

1964 TGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATT 2013
     ||   ||||||   |||   |||      :::   ::::::|||:::
 554 alLysCysGlyArgLeuTyrCys......PheProAsnSerProGluAsn 568

2014 GGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCG 2063
        :::                  |||   ::::::|||::::::
 569 LysAsn.................ProCysAsnIleTyrTyrSer..... 577

2064 GGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACGGATG 2113
                                                :::|||   |||
 578 ................................ProAsnAspGluAspL 583

2114 TGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTA 2163
       |||   |||      ||||||   |||:::            |||
 583 ysGlyMetValLeuProGlyThrLysCysAlaAspArgLysAlaCysSer 599

2164 GATCGGAAGTGCCTACAAATT 2184
     :::   :::|||:::::::::
 600 AsnGlyGlnCysValAspVal 606
```

seq_name: pir2:S55270

seq_documentation_block:
catrocollastatin precursor - western diamondback rattlesnake
C;Species: Crotalus atrox (western diamondback rattlesnake)
C;Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 21-Jul-2000
C;Accession: S55270
R;Zhou, Q.; Smith, J.B.; Grossman, M.H.
Biochem. J. 307, 411-417, 1995
A;Title: Molecular cloning and expression of catrocollastatin, a snake-venom protein
A;Reference number: S55264; MUID:95251603
A;Accession: S55270
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-609 <ZHO>
A;Cross-references: GB:U21003; NID:g710353; PIDN:AAC59672.1; PID:g710354
C;Superfamily: mouse meltrin alpha; disintegrin homology

alignment_scores:
Quality: 723.50 Length: 658
Ratio: 1.874 Gaps: 25
Percent Similarity: 58.663 Percent Identity: 28.267

alignment_block:
US-09-634-252A-2 x S55270 ..

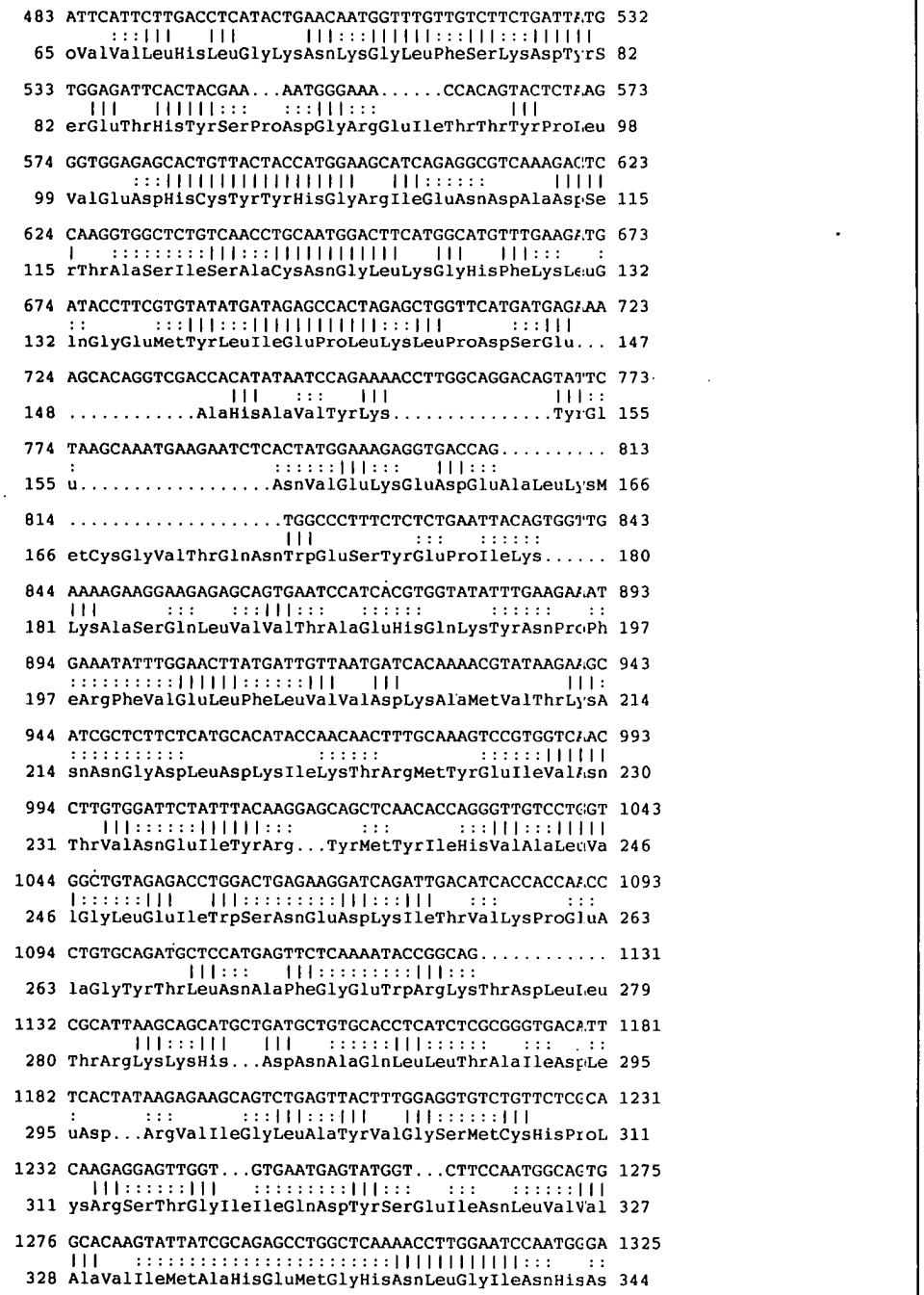
Align seg 1/1 to: S55270 from: 1 to: 609

```
 283 CAGAATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAAT 332
     |||:::||||||      ::::::   :::|||:::   :::   |||::
  17 GlnGlySerSerIleIleLeuGluSerGlyAsnValAsnAspTyrGluVa 33

 333 CACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGAAAGCC 382
     :      |||   ::::::
  33 lIleTyrProArgLysValThrAlaLeu..................... 42

 383 CTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAG 432
                          |||:::      |||   |||::::::
  43 ................ProLysGlyAlaValGlnProLysTyrGluAsp 53

 433 GCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAA 482
     |||:::                :::::::::::::::::   |||:::
  54 AlaMet..............GlnTyrGluLeuLysValAsnGlyGluPr 65
```



seq_name: pir2:I48100

```
seq_documentation_block:
ADAM 5 protein precursor - guinea pig
C;Species: Cavia porcellus (guinea pig)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 04-Feb-2000
C;Accession: I48100
R;Wolfsberg, T.G.; Straight, P.D.; Gerena, R.L.; Huovila, A.P.; Primakoff, P.; Myles, D.
Dev. Biol. 169, 378-383, 1995
A;Title: ADAM, a widely distributed and developmentally regulated gene family encoding m
A;Reference number: I48100; MUID:95269891
A;Accession: I48100
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-777 <RES>
A;Cross-references: EMBL:U22060; NID:g965005; PIDN:AAA74918.1; PID:g965006
C;Genetics:
A;Gene: ADAM5
C;Superfamily: mouse meltrin alpha; disintegrin homology
F;395-480/Domain: disintegrin homology <DIS>


    ment_scores:
           Quality:  687.00            Length:    733
             Ratio:  1.651               Gaps:     29
 Percent Similarity:  56.753   Percent Identity: 27.149

alignment_block:
 US-09-634-252A-2 x I48100   ..

 Align seg 1/1  to: I48100  from: 1  to: 777

     382 CCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAA 431
          |||   ::::::   ::::::   :::|||::::::  :::||||||
      30 ProGluArgIleSerSerValAspThrArgArgHisLeuGluHisAsnVa 46

     432 GGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCA 481
          |||                  ::::::|||       |||::::
      46 lAla.....................TyrAsnIleThrLeuLysGlyLysS 56

     482 AATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTAT 531
          ::::::::::   |||     ::::::::::|||||||||
      56 erTyrValValArgLeu...LysLysGluSerPheLeuSerSerGlySer 71

     532 GTGGAGATTCACTACGAAAATGGGAAACCACAGTACTCTAAG........ 573
          |||         |||:::|||         |||   |||:::
      72 ValIleTyrPheTyrAspAsnArgGlyValGlnArgSerGlnProLeuLe 88

     574 .GGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACT 622
                    |||   |||   |||   :::   |||         |
      88 uProGluMetAspCysSerTyrSerGlyTyrValAlaGlyPheProHisS 105

     623 CCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGAT 672
          ||:::|||::::::::::||||||   ||||||:::|||:::::::::
     105 erArgValValPheAlaThrCysLeuGlyLeuArgGlyValIleGlnPhe 121

     673 GATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAA 722
          ::::::      |||   |||||||||||||||:::::
     122 GluAsnValSerTyrAlaIleGluProLeuGluValLeu........... 134

     723 AAGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCAGGACAGTATT 772
              :::|||      |||:::|||                 |||:
     135 ....SerGlyPheThrHisMetIle.....................TyrG 143

     773 CTAAGCAAATGAAGAATCTCACTATGGAA.........AGAGGTGACCAG 813
          ::::::::      |||      :::           ::::::::::
     143 luGluAsnAsnAspAsnThrHisValProLeuPheGlyLysAsnAsnSer 159

     814 TGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGT 863
          :::       :::   :::|||:::       ::::::   :::|||:::||
     160 TyrAlaArgIleHisAsnLeuGlu......SerGlnGlyArgArgSerVa 174

     864 GAATCCATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTTATGA 913
          |:::   :::                :::   :::|||:::::::::   |

     174 lHisLysThrThrValSerLysLeuSerProArgTyrIleAspMetTyrI 191

     914 TTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACAT 963
          ||||   :::   :::   :::             |||:::
     191 leValValAsnLysAsnLeuPheAspTyrLeuGlySerAspIleLysThr 207

     964 ACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAA 1013
          :::::::::          :::|||::::::||||||:::::::::::
     208 ValThrGlnLysIleIleGlnValIleGlyLeuValAsnAlaMetPhe.. 223

    1014 GGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTG 1063
              ||||||:::   :::|||:::::::   :::::::|||   |||::::
     224 .ThrGlnLeuLysLeuHisValLeuIleSerSerIleGluIleTrpSerA 240

    1064 AGAAGGATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAG 1113
          ::::::::::::::       |||   :::|||   :::   |||   :::
     240 rgSerAsnLysValThrAsnThrArgArgProAspAspAspLeuPheArg 256

    1114 TTCTCAAAATACCGGCAGCGC.........ATTAAGCAGCATGCTGATGC 1154
          ||||||   ::::::::::::         :::|||:::|||
     257 PheSerAspTrpLysArgLysHisValSerLeuLysSerHis........ 270

    1155 TGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTG. 1203
              ::::::         :::||||||             :::|||:::
     271 ....TyrValAlaTyrLeuLeuThrPheAspLysTyrProGluSerIleG 286

    1204 .....AGTTACTTTGGAGGTGTCTGTTCTCGC......ACAAGAGGAGTT 1242
              ::::::      :::::::|||:::::::      :::   |||:::
     286 lyAlaThrPheProGluAsnIleCysAsnGluGluTyrAlaSerGlyIle 302

    1243 GGTGTGAATGAGTATGGTCTTCCAATG...GCAGTGGCACAAGTATTATC 1289
          :::|||         ||||||   :::   :::   |||   ::::::
     303 AlaValTyrProAlaGlyLeuSerLeuGluSerPheAlaValIleIleVa 319

    1290 GCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAA 1339
          |||   |||:::   :::   |||:::::::::::::   |||:::
     319 lGlnLeuLeuSerLeuSerAlaGlyValMetTyrAspThrSerAsp.... 334

    1340 AGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAA 1389
              :::|||   |||       |||           |||   |||   :::
     335 .....SerCysTyrCys......SerThrAspValCysThrMetThrGln 347

    1390 ACAGGGGTG......TCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTT 1433
          :::|||      |||      :::   ||||||   |||||||:::
     348 GluAlaValPheAlaSerGlyLeuLysAspPheSerThrCysSerMetAs 364

    1434 GGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACA 1483
          :::::::::   |||   ::::::   |||   :::||||||   |||
     364 pAsnPheLysTyrPheAlaSerGlnTyrGlyLeuThrCysLeuArgAsnT 381

    1484 GGCCA..............ACAAAGCTATTTGAGCCCACGGAA...... 1512
                              |||   |||   |||   :::
     381 hrSerTyrAspMetProIleTyrLysGlnPheProProArgArgArgArg 397

    1513 ...TGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTT 1559
              |||||||||:::      |||   ||||||||||||||||||||
     398 IleCysGlyAsnSerIleArgGluGluGlyGluGluCysAspCysGlyTh 414

    1560 TCATGTGGAATGC...TATGGATTATGCTGT......AAGAAATGTTCCC 1600
              :::|||   :::      ||||||         :::|||   :
     414 rLeuArgAsnCysThrHisLysLysCysCysAspProMetGlnCysArgM 431

    1601 TCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCA 1650
          :::::::::||||||   |||:::   ||||||||||||   :::::::::
     431 etLysLysGlyAlaLysCysGlyThrGlyProCysCys...ThrValAsp 446

    1651 TGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAAC...GA 1697
          |||  ||||||      :::      ||||||   :::|||:::   ::
     447 CysGlnPheGlnLysAlaAsnValLeuCysArgLysSerValAspLysAs 463

    1698 GTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAA 1747
          :||||||:::   ||||||||||:::|||   ||||||:::|||      :
     463 pCysAspPheAspGluTyrCysAsnGlyArgSerGlyAspCysValHisA 480
```

```
 1748 ATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGC 1797
      ::   :::   |||:::|||:::   |||:::::::::   :::   |||
  480 spThrTyrAlaGlnAsnGlyHisPheCysAspSerGlyGlyAlaPheCys 496

 1798 TACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGG 1847
      :::|||||||:::|||:::|||:::|||:::|||||||||   :::   ||
  497 PheAsnGlyArgCysArgThrHisAspArgGlnCysGlnAlaLeuIleGl 513

 1848 AACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAG 1897
      |     :::   |||:::         ||||||::::::::::|||:::
  513 yGlyAspSerArgGlyAlaProPheAlaCysTyrAspGluValAsnSerA 530

 1898 AAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAG 1947
      ::|||         ||||||||||||:::                  |||
  530 rgGlyAspValTyrGlyAsnCysGlyArgHis.............Gln 541

 1948 TGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACC..... 1989
      |||      :::   ::::::||||||   |||:::||||||
  542 CysTyrIleGlnHisAlaLeuCysGlyLysLeuValCysThrTrpProHi 558

 1990 ....AATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATC. 2034
          :::|||:::         :::::::::   :::   :::::::::
  558 sLysGlnLeuValSerArgValAsnLeuSerValValTyrAlaHisValA 575

 2035 .............ATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATT 2070
                     :::   |||:::         :::      :::
  575 rgAspAspIleCysValAlaThrThrLysThrValArgLysIleIleArg 591

 2071 GACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTA 2120
      |||   |||   :::   |||:::|||   :::|||   |||      ::
  592 AspLeuSerLeuThrThrValLeuLeuProGluAspArgAspGluThrPh 608

 2121 TGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGA 2170
      :|||||||||||||||   |||||||||:::::   |||
  608 eValGluAspGlyThrIleCysGlyProGlyGlnTyrCysAspLysTrpP 625

 2171 AGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCC 2220
        |||   :::::::|||   :::|||   :::||||||      :::
  625 heCysLysGluValGlnPheIleAsnAsnGlySerCysAsnAlaGluIle 641

 2221 AAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTG 2270
                 |||:::|||:::|||:::|||:::|||      :::||
  642 His.........CysGlnGlyArgGlyIleCysAsnAsnLeuAspAsnCy 655

 2271 CATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAG 2320
      |   |||
  655 sHisCysHis..................................... 658

 2321 TTAGGAACCTTCACCCCCCCAAGGATGAAGGACCCAAGGGTCCTAGTGCC 2370
        :::::::::   ||||||:::            ||||||
  659 ..LysGlyPheValProProGluCysAlaProLysLysGlyGln...... 672

 2371 ACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCTAT 2420
                 :::|||||:::      :::   ||||||   :::
  673 ............PheGlySerLeuAspAspGlyHisLeuValGluThrTh 685

 2421 TGTCCTTGGGGGCACAGGCTGGGGATTTAAAAATGTCAAGAAGAGAAGG 2469
            :::           ||||||:::::::::::   ||||||
  685 rLysThrSer............GlyPheArgLysIleAsnMetArgArg 697
```

```
seq_name: pir2:A60385

seq_documentation_block:
monocyte surface antigen MS2 precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 03-Feb-1993 #sequence_revision 03-Feb-1993 #text_change 26-Aug-1999
C;Accession: A60385
R;Yoshida, S.; Setoguchi, M.; Higuchi, Y.; Akizuki, S.; Yamamoto, S.
Int. Immunol. 2, 585-591, 1990
A;Title: Molecular cloning of cDNA encoding MS2 antigen, a novel cell surface antigen st
A;Reference number: A60385; MUID:91197896
A;Accession: A60385
A;Molecule type: mRNA
A;Residues: 1-826 <YOS>
A;Cross-references: EMBL:X13335
C;Superfamily: mouse meltrin alpha; disintegrin homology
C;Keywords: glycoprotein; surface antigen; transmembrane protein
F;1-14/Domain: signal sequence #status predicted <SIG>
F;402-484/Domain: disintegrin homology <DIS>
F;659-683/Domain: transmembrane #status predicted <TMM>
F;330/Active site: Glu #status predicted


alignment_scores:
            Quality:  685.50              Length:    733
              Ratio:  1.731                 Gaps:     29
 Percent Similarity:  54.025   Percent Identity: 29.332

alignment_block:
 US-09-634-252A-2 x A60385   ..

 Align seg 1/1  to: A60385  from: 1  to: 826

     322 CAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACATCAACCAAGA 371
         |||   |||::::::   |||   ||||||
      26 GlnTyrGluValValTrpProArgArgLeu..................... 35

     372 CTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAA 421
                                       :::::::::|||:::
      36 .....................AlaAlaSerArgSerArgArgAlaLeuP 45

     422 AACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCC 471
             :::   :::   :::      :::   |||:::   :::   :::
      45 roSerHisTrpGlyGlnTyrProGluSerLeuSerTyrAlaLeuGlyThr 61

     472 TTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTC 521
         |||       |||   |||   |||       |||:::   ||||||::
      62 SerGlyHisValPheThrLeuHisLeuArgLysAsnArgAspLeuLeuGl 78

     522 TTCTGATTATGTGGAGATTCAC...TACGAAAATGGGAAA......CCAC 562
         :|||:::|||:::|||   :::          ||||||:::         :
      78 ySerSerTyrThrGluThrTyrSerAlaAlaAsnGlySerGluValThrG 95

     563 AGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGC 612
         ::       :::       :::||||||   |||:::|||   ::::::|||
      95 luGlnLeuGlnGluGlnAspHisCysLeuTyrGlnGlyHisValGluGly 111

     613 GTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCAT 662
         :::   |||   :::::::::|||||||||   ||||||:::|||::
     112 TyrGluGlySerAlaAlaSerIleSerThrCysAlaGlyLeuArgGlyPh 128

     663 GTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTC 712
         :|||:::      :::   |||::::::|||||||||||||:::
     128 ePheArgValGlySerThrValHisLeuIleGluProLeuAspAla.... 143

     713 ATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCA 762
           ||||||:::
     144 ..AspGluGlu..................................... 146

     763 GGACAGTATTCTAAG...CAAATGAAGAATCTCACTATGGAAAGAGGT.. 807
         ||||||::::::      |||   |||:::|||   ::::::   |||
     147 GlyGlnHisAlaMetTyrGlnAlaLysHisLeuGlnGlnLysAlaGlyTh 163

     808 ..........................GACCAGTGGCCCTTTCTCTCTG 829
                                    |||      |||         |
     163 rCysGlyValLysAspThrAsnLeuAsnAspLeuGlyProArgAlaLeuG 180

     830 AATTA.................CAGTGGTTGAAAAGAAGGAAGAGAGCA 861
         ||:::                ::::||||||
     180 luIleTyrArgAlaGlnProArgAsnTrpLeu.................. 190

     862 GTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTTAT 911
                           |||   :::|||   :::|||:::||||||
     191 .................IleProArgGluThrArgTyrValGluLeuTy 201
```

```
 912 GATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCAC 961
 201 rValValAlaAspSerGlnGluPheGlnLysLeuGlySerArgGluAla. 217

 962 ATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTAC 1011
 218 ..ValArgGlnArgValLeuGluValValAsnHisValAspLysLeuTyr 233

1012 AAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGAC 1061
 234 GlnGlu...LeuSerPheArgValValLeuValGlyLeuGluIleTrp.. 248

1062 TGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATG 1111
 249 .AsnLysAspLysPheTyrIleSerArgTyrAlaAsnValThrLeuGluA 265

1112 AGTTCTCAAAATACCGGCAGCGC......ATTAAGCAGCATGCT...GAT 1152
 265 snPheLeuSerTrpArgGluGlnAsnLeuGlnGlyGlnHisProHisAsp 281

1153 GCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCT 1202
 282 AsnValGlnLeuIleThrGlyValAspPheIleGlySerThrValGlyLe 298

1203 GAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGA...GTTGGTGTGA 1249
 298 uAlaLysValSerAlaLeuCysSerArgHisSerGlyAlaValAsnGlnA 315

1250 ATGAGTATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTG 1299
 315 spHisSerLysAsnSerIleGlyValAlaSerThrMetAlaHisGluLeu 331

1300 GCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATG 1349
 332 GlyHisAsnLeuGlyMetSerHisAspGluAspIle......ProGlyCy 346

1350 TGACTGCACAGAA......TCCTGGGGTGGCTGCATCATGGAGGAAACA. 1392
 346 sTyrCysProGluProArgGluGlyGlyGlyCysIleMetThrGluSerI 363

1393 ..GGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTAT 1440
 363 leGlySerLysPheProArgIlePheSerArgCysSerLysIleAspLeu 379

1441 AGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCA.. 1488
 380 GluSerPheValThrLysProGlnThrGlyCysLeuThrAsnValProAs 396

1489 ....ACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATACGTGGAAG 1534
 396 pValAsnArgPheValGlyGlyProValCysGlyAsnLeuPheValGluH 413

 535 CTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTATGGATTATGC 1584
 413 isGlyGluGlnCysAspCysGlyThrProGlnAspCysGlnAsnProCys 429

1585 TGTAAG......AAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGG 1628
 430 CysAsnAlaThrThrCysGlnLeuValLysGlyAlaGluCysAlaSerGl 446

1629 GCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAAT 1678
 446 yThrCysCysHis......GluCysLysValLysProAlaGlyGluValC 461

1679 GCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTACTGGAGAC 1728
 461 ysArgLeuSerLysAspLysCysAspLeuGluGluPheCysAspGlyArg 477

1729 TCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATGCATGCAA 1778
 478 LysProThrCysProGluAspAlaPheGlnGlnAsnGlyThrProCys.. 493

1779 TCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAACC 1828
```

```
 494 ....ProGlyGlyTyrCysPheAspGlySerCysProThrLeuAlaGlnG 509

1829 AGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGC 1878
 509 lnCysArgAspLeuTrpGlyProGlyAlaArgValAlaAlaAspSerCys 525

1879 TATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGA 1928
 526 Tyr.....ThrPheSerIleProProGlyCysAsnGlyArg.. 537

1929 TGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCT 1978
 538 .....MetTyrSerGlyArgIleAsnArgCysGlyAlaL 549

1979 TACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGT 2028
 549 euTyrCysGluGly.....GlyGlnLysProLeu 558

2029 GAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAG 2078
 559 GluArgSerPheCysThrPheSerSerAsnHisGlyValCysHisAlaLe 575

2079 TGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTATGTAGAAG 2128
 575 uGlyThrGlySerAsnIleAsp.....ThrPheGluLeuValLeuG 589

2129 ATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTA 2178
 589 lnGlyThrLysCysGluGluGlyLysValCysMetAspGlySerCysGln 605

2179 CAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAA 2228
 606 AspLeuArgValTyrArgSerGluAsnCys.....SerAl 617

2229 AGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTG 2278
 617 aLysCysAsnAsnHisGlyValCysAsnHisLysArgGluCysHisCysH 634

2279 ATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGTTAGGAAC 2328
 634 isLysGlyTrpAlaProProAsnCys.....ValGlnArg 645

2329 CTTCACCCCCCCAAGGATGAAGGACCCAAGGGTCCTAGTGCCACCAATCT 2378
 646 LeuAlaAspValSerAspGluGlnAlaAlaSerThrSerLeuProValSe 662

2379 CATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCTATTGTCCTT 2427
 662 rValVal.....ValValLeuValIleLeuValAlaAlaMetValIle 676
```

```
seq_name: pir2:S47645

seq_documentation_block:
tMDC I protein - crab-eating macaque
C;Species: Macaca fascicularis (crab-eating macaque)
C;Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 21-Jul-2000
C;Accession: S47645
R;Barker, H.L.; Perry, A.C.F.; Jones, R.; Hall, L.
Biochim. Biophys. Acta 1218, 429-431, 1994
A;Title: Sequence and expression of a monkey testicular transcript encoding tMDC I, a
A;Reference number: S47645; MUID:94325353
A;Accession: S47645
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-736 <BAR>
A;Cross-references: EMBL:X76637; NID:g535016; PIDN:CAA54085.1; PID:g535017
C;Superfamily: mouse meltrin alpha; disintegrin homology
F;392-477/Domain: disintegrin homology <DIS>


alignment_scores:
          Quality:  685.00            Length:    763
```

```
           Ratio:  1.635              Gaps:     31
 Percent Similarity:  54.915   Percent Identity: 26.212

alignment_block:
 US-09-634-252A-2 x S47645   ..

 Align seg 1/1  to: S47645  from: 1  to: 736

     328 GAAATCACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCCGA 377
         :::||||||:::|||   ::::::          |||:::   :::
      27 GlnIleThrValProArgLysIleGlnThrAsnThrAsnAspGlyAspVa 43

     378 AAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATA 427
          |||     |||
      43 lSerGluThrHis..................................... 47

     428 ATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGC 477
                               |||::::::::::|||:::      |||
      48 .....................AlaThrTyrSerIleLysIleGluGly 56

     478 TCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGA 527
         :::   :::   :::   |||      ::::::   |||
      57 LysProTyrThrPheLeuLeuGluLysGlnSerPheLeuHisProHisPh 73

     528 TTATGTGGAGATTCACTACGAAAATGGG............AAACCACAGT 565
             |||   ::::::   |||:::|||                ::::
      73 eLeuValTyrLeuTyrAsnGluSerGlyThrLeuTyrValAspSerSerP 90

     566 ACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGCGTC 515
         ::|||||||||       ||||||:::|||:::|||   :::       :::
      90 heSerLysGly......HisCysPheTyrGlnGlyTyrValAlaAspIle 104

     616 AAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTT 565
             |||   |||:::|||   ||||||:::||||||:::|||::::::
     105 ProLysSerAlaValThrLeuArgThrCysSerGlyLeuArgGlyLeuLe 121

     666 TGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATG 715
         ::::   |||::::::   |||   ||||||||||
     121 uGlnLeuAspAsnIleSerTyrGlyIleGluProLeu............. 133

     716 ATGAGAAAAGCACAGGTCGACCACATATA.........ATCCAGAAAACC 756
           |||:::|||             |||:::         |||::::::
     134 ..GluSerSerProThrTyrGluHisValValTyrArgIleLysAsnAsp 149

     757 TTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAGG 806
             |||::::::|||   |||
     150 AlaIleGlyHisPheSerPheGln........................... 157

     807 TGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAAGA 856
          ::::::::::|||                |||::::::   :::      |
     158 .GluAsnTyrProVal.........AlaGlnTyrIleAspGlnSerTyrA 171

     857 GAGCAGTGAATCCATCA.........CGTGGTATATTTGAAGAAATGAAA 897
         ||   :::      |||            |||   :::   :::   :::
     171 rgIleLeuValLysSerAspIleAsnSerGlyAlaMetLeuSerLysArg 187

     898 TATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCG 947
            |||::::::::::|||:::   |||         |||
     188 ThrLeuLysIleGlnIleIleMetAspLysAlaMetTyrAspTyrMetGl 204

     948 CTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTG 997
         |||:::   |||   :::   :::   :::   ::::::   :::||||:
     204 ySerAlaValAlaValAlaValGluLysValPheGlnIlePheGlyLeuI 221

     998 TGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCT 1047
         ::::::::::::::   :::|||||||||       |||:::|||   :::
     221 leAsnThrMetPhe...SerGlnLeuAsnIleThrValMetLeuSerSer 236

    1048 GTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGT 1097
         :::|||   |||::::::::::::::|||:::   :::   :::
     237 LeuGluIleTrpSerAspGlnAsnLysIleSerThrSerGlyHisAlaAs 253

    1098 GCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATT.........A 1138

      :::::::|||:::::|||      :::::::|||::::::        :
  253 pGluIleLeuGlnArgPheLeuLeuTrpLysGlnLysLeuLeuPheGlnA 270

 1139 AGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTAT 1188
      :::::|||      :::      ||||||   |||      |||
  270 rgSerHisAspMetThrTyrLeuLeuIleTyrArg......AsnHisSer 284

 1189 AAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGG 1238
             :::   :::|||   |||   :::|||:::         ::
  285 ThrTyrValGlyAlaThrTyrHisGlyMetAlaCysAspProLysPheAl 301

 1239 AGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTG.........GCAC 1279
      ::::|||:::      |||         ::::::|||         :::
  301 aThrGlyIleAlaLeuTyrProLysLysIleThrValGluAlaPheSerV 318

 1280 AAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCT 1329
        |||::::::|||   |||:::   |||||||||:::   ::::::
  318 alValMetAlaGlnLeuLeuGlyIleAsnLeuGlyLeuThrTyrAspAsp 334

 1330 TCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCAT 1379
                  :::|||   |||                    |||||
  335 IleTyr.........AsnCysTyrCysProGlyPro......ThrCysIl 346

 1380 CATGGAGGAAACAGGGGTG...TCCCATTCTCGAAAA...TTTTCAAAGT 1423
      ||||:::      ::::::   ||||||:::   |||   ||||||:::|
  346 eMetAsnProAspAlaIleArgSerHisGlyMetLysPhePheSerSerC 363

 1424 GCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGC 1473
      |||||:::   |||:::::::::::::::   :::            |||
  363 ysSerMetAspGluPheLysGlnIleValLeuGlnProGluLeuLysCys 379

 1474 CTTTTCAACAGGCCAACAAAGCTATTT.........GAGCCCACGGAATG 1514
      :::   ::::::      :::   :::          :::   :::   ||
  380 PheGlnAspLysThrIleSerGluMetThrTyrGlnLysSerAlaThrCy 396

 1515 TGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATG 1564
      ||||||||||   :::|||      |||:::||||||||||||:::::
  396 sGlyAsnGlyIleLeuGluProThrGluGlnCysAspCysGlyTyrArgL 413

 1565 TGGAATGC...TATGGATTATGCTGTAAG......AAATGTTCCCTCTCC 1605
           |||   :::      |||||:::          |||:::|||
  413 ysAlaCysThrPheLysLysCysCysAsnProAlaAspCysThrLeuIle 429

 1606 AACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCT 1655
      :::   |||:::|||::::::|||||||||||||||||||'  :::|||
  430 GlyPheAlaGluCysGlySerGlyProCysCysAsnAsnLysThrCysTh 446

 1656 TTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATA 1705
        :::      ||||||:::   ||||||   :::::::::   ||||||:
  446 rIleTyrAlaArgGlyHisValCysArgLysSerIleAspMetCysAspP 463

 1706 TTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCAT 1755
      ::   |||||||||:::|||   |||      |||   |||::::::
  463 heProGluTyrCysAsnGlyThrSerGluPheCysValProAspValLys 479

 1756 AAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGG 1805
           |||         |||:::::::::   :::   |||::::::||
  480 AlaAlaAspLeuGluTyrCysSerAsnLysThrSerTyrCysPheLysGl 496

 1806 CGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGG 1855
      |   |||:::   ||||||:::||||||:::   ::::::|||      :
  496 yIleCysArgAspArgAspArgGlnCysSerGlnLeuPheGlyLysPheS 513

 1856 CTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACT 1905
      ::   :::::::::   :::|||   |||:::::|||   ::::::
  513 erLysSerAlaAsnLeuLeuCysThrGluGluValAsnPheGlnAsnAsp 529

 1906 GAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAA 1955
      :::   ||||||||||||:::                    :::|||
  530 LysPheGlyAsnCysGlyGln.................ArgCysProPh 540

 1956 ACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTC 2005
        |||::::::||||||   ::::::|||
```

```
    540 ePheAspIleLeuCysGlyLysIleValCys..................  550

   2006 CACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCC.........  2046
                :::::::::|||:::::|||   :::
    551 ........HisTrpIleHisSerGluLeuValProMetThrAspLeuAsp  564

   2047 .........TTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCA 2087
                  |||      |||:::|||      |||   :::|||||
    565 IleGlnTyrThrTyrLeuGlyGlyHisVal......CysLeuSerAlaHi  579

   2088 TGTAGTTTTAGATGATGATACGGATGTGGGC...TATGTAGAAGATGGAA 2134
        |   :::   :::   :::   ::::::|||   |||:::||||||   |
    579 s...AlaArgAsnAlaSerLysGlnLeuGlyThrTyrThrGluAspIleT  595

   2135 CGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAATT 2184
        ||   ||||||   :::   :::|||   :::|||   |||      :::
    595 hrAlaCysGlyGlnGlnLysValCysArgSerArgGlyCysArgTyrLeu  611

   2185 CAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTG 2234
        :::   ||||||:::::::::|||      :::                ||
    612 SerGluLeuAsnIleThrLysCysThrThrAsn..............Cy  623

    235 TTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCA 2284
        |:::   :::|||:::|||:::::::::      |||   ||||||
    623 sGlyGlnAsnGlyIleCysAsnAspArgPheHisCysGlnCysAspProG  640

   2285 CCTGGGCAGGGACAGATTGC......AGTATCCGGGATCCAGTTAGGAAC 2328
          :::|||      ||||||      |||:::   :::|||      :::
    640 lyTyrAlaProProAspCysAspProSerMetSerSerProGlyGlySer  656

   2329 CTTCACCCCCCCAAGGATGAAGGA........................... 2352
        :::            |||:::|||
    657 Ile............AspAspGlyPheTrpLeuSerValAspLysSerVa  669

   2353 ..........CCCAAG.........GGTCCTAGTGCCACCAATCTCATAA 2383
                  ||||||         :::|||:::   ::::::|||:::|
    669 lProLeuPheProLysGlnArgAlaAlaProLysAsnAsnGlyLeuLeuI  686

   2384 TAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCTATTGTCCTTGGGGGC 2433
        ||:::             :::   |||:::      |||::::::   :::
    686 leSerPheTyrIlePheSerProLeuLeuIleLeuIleAlaIleValSer  702

   2434 ACAGGCTGGGGATTTAAAAATGTCAAGAAGAGAAGGTTC 2472
              |||:::            |||   :::|||||
    703 LeuLysTrpAsn............LysMetLysArgPhe 711

seq_name: pir2:S47656

seq_documentation_block:
tMDC II protein - crab-eating macaque
C;Species: Macaca fascicularis (crab-eating macaque)
[redacted]e: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 21-Jul-2000
[redacted]ession: S47656
[redacted]ry, A.C.F.; Barker, H.L.; Jones, R.; Hall, L.
Biochim. Biophys. Acta 1207, 134-137, 1994
A;Title: Genetic evidence for an additional member of the metalloproteinase-like, disint
A;Reference number: S47656; MUID:94318664
A;Accession: S47656
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-756 <PER>
A;Cross-references: EMBL:X77619; NID:g531477; PIDN:CAA54713.1; PID:g531478
C;Superfamily: mouse meltrin alpha; disintegrin homology
F;388-473/Domain: disintegrin homology <DIS>


alignment_scores:
            Quality:  681.50            Length:    796
              Ratio:  1.619              Gaps:     36
 Percent Similarity:  52.889   Percent Identity: 26.884

alignment_block:
 US-09-634-252A-2 x S47656   ..
```

```
Align seg 1/1  to: S47656  from: 1  to: 756

   136 CTGCTTCTCGTCCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCC 185
       ::::::|||:::|||:::|||||      |||:::
     1 MetPheLeuLeuLeuValLeuLeuThrGlyLeuGly.............  12

   186 CCGCGCCTGGGGGGCTGCTGCGCCCAGCGCTCCGCATTGGAATGAAACTG 235
                |||      |||         ||||||
    13 ..........GlyMetHisAlaAspLeuAsnProHis.............  21

   236 CAGAAAAAAATTTGGGAGTCCTGGCAGATGAAGACAATACATTGCAACAG 285
                                        :::   ||||||
    22 ............................LysThrPheLeuGlnThr  27

   286 AATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAATCAC 335
       :::      ::::::|||||   |||:::|||              ||
    28 ThrIleProGluLysIleSerSerSerAspAlaLys...........Th  40

   336 ACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGAAAGCCCTT 385
       |   |||::::::   :::   |||:::         :::   :::||||
    40 rAspProGluHisAsnValValTyrMetIleThrIleGluGlyLysProT  57

   386 ATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAGGCT 435
       ||   |||                  |||   :::|||::::::
    57 yrPheVal..................HisLeuLysLysGlnSer......  65

   436 GTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATT 485
             |||::::::|||||||||
    66 ...IleLeuSerSerAlaSerPhe.........................  72

   486 CATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGG 535

    72 ................................................  72

   536 AGATTCAC...TACGAAAATGGGAAACCACAGTACTCTAAGGGT...... 576
         ||||||   |||:::::::::      ::::::||||||
    73 ..IleHisSerTyrAspLysAsnAspIleArgHisSerLysProLeuLeu  88

   577 ...GGAGAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTC 623
                     |||   |||:::|||   :::   |||:::   :::||
    89 ValGlnMetAspCysAsnTyrAsnGlyTyrValAlaGlyIleProAsnSe  105

   624 CAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATG 673
       |   |||:::||||||:::|||:::||||||:::|||   ::::::
   105 rLeuValThrLeuSerValCysSerGlyLeuArgGlyThrMetGlnLeuL  122

   674 ATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAA 723
         ::::::   |||   |||||||||:::|||   |||
   122 ysAsnIleSerTyrGlyIleGluProMetGluAlaVal............  134

   724 AGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCAGGACAGTATTC 773
          ::|||      |||   |||                        |||::
   135 ...SerGlyPheIleHisLysIle.....................TyrGl  143

   774 TAAGCAAATG......AAGAATCTCACTATGGAAAGAGGTGACCAGTGGC 817
       ::::::::::         |||:::   :::|||:::::::|||   :::
   143 uGluLysPheAlaAspThrAsnIleLeuLeuGluGluAsnAspThrTyrS  160

   818 CC...TTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTG 864
             |||   ||||||   |||      |||      :::::::::
   160 erTrpPheAsnSerGluTyrGlnValArgLysSerSerGluLysThr...  175

   865 AATCCATCACGTGGTATATTTGAAGAAATG......AAATATTTGGAACT 908
                         |||   ::::::      :::|||:::|||::
   176 ..............AspPheIleLysLeuProProArgTyrIleGluMe  187

   909 TATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATG 958
       :   ||||||   |||   :::   :::            |||:::
   187 tHisIleValValAspLysAsnLeuPheAspTyrMetGlySerAspIleA  204

   959 CACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATT 1008
             :::::::::   :::   ::::::::::::|||||:::::::::
```

```
 204 snAlaValThrGlnLysValIleGlnIleIleGlyLeuValAsnThrMet 220

1009 TACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTG 1058
            ||||||:::      |||:::::::   :::::::|||   ||
 221 Leu...ThrGlnLeuGlnLeuThrValIleIleSerSerIleGluIleTr 236

1059 GACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCC 1108
     |:::::::|||::::::|||:::   |||   :::          :::|||
 236 pSerAsnLysAsnLysIleSerThrThrGlyHisAlaGluTyrValLeuL 253

1109 ATGAGTTCTCAAAATACCGGCAGCGCATTAAGCAGCATGCTGAT...GCT 1155
       ||||||   :::::::::          |||:::|||   :::
 253 euGluPhePheGluTrpLys.........LysAspHisLeuAsnPheLys 266

1156 GTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTG.. 1203
        |||   |||:::   :::       :::   |||       :::|||
 267 ProHisGlnIleAlaTyrLeuPheValTyrArgLysLeuProThrLeuIl 283

1204 .......AGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGTG 1246
             ::::::   |||   ||||||::::::      :::::::::|
 283 eGlyAlaThrPheProGlyGlnValCysAsnLysAspPheAlaAlaAlaV 300

1247 TGAATGAGTAT......GGTCTTCCAATG...GCAGTGGCACAAGTATTA 1287
     ||       |||      ||||||   :::   :::   :::   ::::::
 300 alAlaLeuTyrProGluGlyLeuSerLeuGluSerTyrThrValIleIle 316

1288 TCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAG 1337
        |||   |||:::   |||||||||:::   ::::::       ::
 317 ValGlnLeuLeuGlyLeuAsnLeuGlyLeuThrTyrAsp.........Ly 330

1338 AAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGG 1387
     :       |||   |||      |||          |||   |||
 330 sThrAspThrCysHisCys......SerGlyAspValCysThrMetThrP 345

1388 AAACAGGGGTGTCCCATTCT......CGAAAATTTTCAAAGTGCAGCATT 1431
          :::|||      :::      :::   ||||||   ||||||:::
 345 roLysAlaValTyrSerGlyGlyValLysAspPheSerValCysSerLeu 361

1432 TTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAA 1481
        :::::::::   :::   ::::::::::|||   :::||||||   ::
 362 AspAspPheLysTyrIleSerSerHisAsnGlyLeuThrCysLeuGlnTh 378

1482 CAGGCCAACAAAGCTATTTGAGCCCACGGAA.........TGTGGAAATG 1522
     ::::|||·  ::::::       |||:::          |||||||||
 378 rAsnProLeuGluMetProThrTyrThrGlnArgArgIleCysGlyAsnG 395

1523 GATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGC 1572
     ||   :::|||:::|||||||||||||||||||||||          :::|||
 395 lyLeuLeuGluGlyGlyGluGluCysAspCysGlyAsnLysAspAsnCys 411

1573 ...TATGGATTATGCTGTAAG......AAATGTTCCCTCTCCAACGGGGC 1613
        :::   |||||||||             |||   |||:::::::::||
 412 ThrHisLysLeuCysCysAspAlaLeuThrCysArgLeuLysAspAsnAl 428

1614 TCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGC 1663
     |:::|||::::::|||   ||||||::::::   :::|||   |||:::|
 428 aGlnCysGlySerGlyAspCysCysSerLys...AspCysLysPheLysP 444

1664 CACGAGGGTATGAATGCCGGGATGCTGTGAAC...GAGTGTGATATTACT 1710
     ||   :::      ||||||   :::|||:::   |||||||||:::|||
 444 roAlaAsnThrIleCysArgLysSerValAspValGluCysAspPheThr 460

1711 GAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCA 1760
     |||:::|||:::|||:::          |||      :::   :::   ::
 461 GluPheCysAsnGlySerTyrProTyrCysLeuLeuAspThrTyrValAr 477

1761 AGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGT 1810
     :||||||      |||:::::::::   :::   |||::::::|||:::|
 477 gAspGlyGluTyrCysAspSerGlyGlyAlaPheCysPheGlnGlyArgC 494

1811 GCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCA 1860
     ||:::|||   |||:::||||||:::   :::   |||      :::
 494 ysArgThrPheAspLysGlnCysAspAspLeuIleGlyArgGlySerArg 510

1861 GGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAA 1910
     |||:::      ||||||||:::::::::||||||:::|||   :::
  511 GlyAlaProIlePheCysTyrAspGluIleAsnThrArgGlyAspLysPh 527

1911 GGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATG 1960
      ||||||||||||   :::                  |||       :::
  527 eGlyAsnCysGlyThrGlu..............TyrCysLeuPheGlnH 539

1961 ATGTGTTCTGTGGATTCTTACTCTGTACC............AATCTTACT 1998
       ::::::||||||   |||:::||||||            :::|||
  539 isIleLeuCysGlyLysLeuValCysThrTrpGluHisLysAspLeuIle 555

1999 CGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATC.............. 2034
           |||:::::::::   :::   :::::::::
  556 SerArgProAsnLeuSerValIleTyrAlaHisValArgAspGlnThrCy 572

2035 .............ATTCCA............................... 2040
                  :::|||
  572 sValSerThrTyrLeuProSerArgLysProProProValAlaSerThrV 589

2041 ........ACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGT 2082
              ||||||:::|||
  589 alSerLysThrSerTyrTyrSerVal........................ 597

2083 GCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTATGTAGAAGATGG 2132
                       ||||||   |||      :::|||:::|||||
  598 ....................AspAspArgAspGluThrPheValGlnAspGl 608

2133 AACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAA 2182
     |:::   |||||||||:::|||   |||:::      :::|||   ::::
  608 ySerValCysGlyProAspMetTyrCysPheLysMetArgCysLysHisV 625

2183 TTCAAGCCCTA...AATATGAGCAGCTGTCCACTCGATTCCAAGGGTAAA 2229
     :::::   |||   :::::::::::::|||       :::   :::
  625 alArgPheLeuMetAspPheGluThrCysGluAlaSerIleGlu...... 639

2230 GTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGA 2279
        |||||||||||||||:::|||:::|||          |||   |||::
  640 ...CysSerGlyHisGlyIleCysAsnAsnPheAsnHisCysHisCysGl 655

2280 TTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGTTAGGAACC 2329
     :       :::             |||:::          |||   ::::::
  655 uLysGlyTyrAsnProProHisCysLys.........ProLysLysGluA 669

2330 TTCACCCCCCCAAGGATGAAGGACCCAAGGGTCCTAGT 2367
                     |||:::|||         |||:::
  669 laPheGlySerThrAspAspGlyHisLeuValProAla 681

seq_name: pir2:T18900

seq_documentation_block:
disintegrin and metalloproteinase (EC 3.4.24.-) - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T18900; T23241
R;McMurray, A.
submitted to the EMBL Data Library, November 1996
A;Reference number: Z19041
A;Accession: T18900
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-952 <WIL>
A;Cross-references: EMBL:Z81460; PIDN:CAB03832.1; GSPDB:GN00028; CESP:C04A11.4
A;Experimental source: clone C04A11
R;Bardill, S.
submitted to the EMBL Data Library, March 1996
A;Reference number: Z19714
A;Accession: T23241
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-952 <WI2>
A;Cross-references: EMBL:Z70209; PIDN:CAA94147.1; GSPDB:GN00028; CESP:C04A11.4
```

```
A;Experimental source: clone K02D3
C;Genetics:
A;Gene: CESP:C04A11.4
A;Map position: X
A;Introns: 66/2; 82/3; 106/1; 139/1; 189/3; 213/3; 269/2; 297/3; 475/3; 533/2; 604/3; 63
C;Superfamily: mouse meltrin alpha; disintegrin homology
C;Keywords: hydrolase; metalloproteinase; zinc
F;378-461/Domain: disintegrin homology <DIS>
F;312,316,322/Binding site: zinc (His) #status predicted
F;313/Active site: Glu #status predicted


alignment_scores:
            Quality:  667.00              Length:    794
              Ratio:  1.548                Gaps:     37
 Percent Similarity:  54.282   Percent Identity: 26.700

alignment_block:
 US-09-634-252A-2 x T18900   ..

gn seg 1/1  to: T18900  from: 1  to: 952

  394 CTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCT 443
      ||||||   |||         :::::::::   |||
    5 LeuAspLeuLysLeuSerSerArgArgGlnTrpAsnPro..... 17

  444 GGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCC.....A 481
       :::::::::      ::::::|||:::   ||||||
   18 .ValArgCysProValArgLeuGluValAspGlySerAlaGlnThrProT 34

  482 AATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTAT 531
       :::::::::      |||||||||         |||   |||
   34 hrSerLeuValGlnThrAlaLeuAsnAsnPro......SerPheAspLeu 48

  532 GTGGAGATTCACTACGAAAATGGGAAA..... 558
      |||   |||         ||||||:::
   49 Val...IleAlaAlaProAsnGlyGlnAsnValTyrIleProPheThrGl 64

  559 .....CCACAGTACTCTA 571
                        |||:::   |||:
   64 uAspArgLysLeuPheThrAlaAsnIleAlaAspAspProSerThrSerS 81

  572 AGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGAC 621
      ::      :::||||||:::::::::|||           :::::|||
   81 erLeuIleSerHisCysHisPheGluGlyVal.....ThrGluAsp 94

  622 TCCAAGGTGGCTCTGTCAACCTGCAAT.....GGACTT.. 654
      ::::::   |||||||||   |||:::           ||||||
   95 GlyArgHisAlaLeuSerLeuCysAspProGlyGluIleThrGlyLeuLe 111

  655 .....CATGGCATGTTTGAAGATGATACCTTCGTGT 685
                      |||:::      :::::::::      :
  111 uMetThrGlnThrAsnArgPheGlyLeuSerThrSerAsnAsnGlySerP 128

  686 ATATGATAGAGCCACTA.....GAGCTGGTT 711
      ::::::::   |||                :::||||||
  128 heValLeuIleProTyrValGluAsnAsnCysAspLeuGlySerLeuVal 144

  712 CATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAACCTTGGC 761
      |||:::::::::         |||
  145 HisSerSerSerArgLysLysArg..... 152

  762 AGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAGGTGACC 811
          |||   :::||||||      |||   :::::::::|||:::   :
  153 ....GlnIleGlyLysGln......AsnThrValIleAspArgAsnProS 166

  812 AGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAAGAGAGCA 861
      :::::                  :::   |||      |||||||||
  166 erTyr...........IleArgGluHisLeuAspGlyArgLysArg... 177

  862 GTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTTAT 911
                              ::::::||||||
  178 .....PheValGluLeuAl 182

  912 GATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCAC 961
       :::|||   |||:::::::::|||   |||:::   |||::::::
  182 aLeuValAlaAspTyrSerValTyrThrLysTyrAspSerAspGluLysL 199

  962 ATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTAC 1011
        :::|||::::::   :::::::::::::|||:::::::::|||:::|||
  199 ysValAsnAspTyrMetGlnGlnThrMetAsnIleLeuAsnSerLeuTyr 215

 1012 AAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGAC 1061
              ||||||   |||::::::||||||      |||   |||
  216 ...PheProLeuAsnIleArgIleThrLeuValHisSerGluIleTrpLy 231

 1062 TGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATG 1111
       :::   |||||||||:::::      :::      :::   |||::::
  231 sLysGlyAspGlnIleSerValIleProAspSerLysGluThrLeuAsnA 248

 1112 AGTTCTCAAAATACCGGCAG...CGCATTAAGCAGCAT...GCTGATGCT 1155
       ::|||   :::|||::::::      :::|||:::|||      |||:::
  248 snPheMetGluTyrLysLysIleMetLeuLysAspHisPhePheAspThr 264

 1156 GTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAG 1205
        :::|||::::::   :::   |||                :::   ::
  265 GlyTyrLeuMetThrThrLeuLysPheAspGluGlyValValGlyLysAl 281

 1206 TTACTTTGGAGGTGTCTGTTCT......CGCACAAGAGGAGTTGGTGTGA 1249
       :|||   |||   :::||||||         :::   |||:::   |||:
  281 aTyrLysGlyThrMetCysSerTyrAspTyrSerGlyGlyIleTyrValA 298

 1250 ATGAGTATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTG 1299
       :::::   :::      :::   ::::::   ::::::::::::::|||
  298 spHisAsnAsnAspThrValGluThrValAlaThrPheAlaHisGluLeu 314

 1300 GCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATG 1349
       ::::::::::|||::::::   :::|||::::::   |||      ||
  315 GlyHisThrPheGlyMetAspHisAspProAsnAsp...LysAspValCy 330

 1350 TGACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGT 1399
       |   |||                  |||||||||:::      :::   :
  330 sTyrCys......ProMetProArgCysIleMetAsnProGlnSer...G 344

 1400 CCCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTT 1449
       ::|||   :::   :::|||:::||||||:::   :::      :::
  344 lyHisMetGluValTrpSerGluCysSerValLysAsnLeuAlaSerGly 360

 1450 TTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATT 1499
       :::::||||||         ||||||||||||:::|||   |||
  361 PheAsnArgGlyIleAspLeuCysLeuPheAsnGluProGlyLysLysPr 377

 1500 TGAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTG 1549
       :::   ::::::||||||||||||   ||||||   ||||||||||||
  377 oSerAspAlaLysCysGlyAsnGlyIleValGluProGlyGluGluCysA 394

 1550 ATTGTGGTTTTCATGTGGAATGCTATGGATTATGCTGT......AAGAAA 1593
       |||||||      ::::::|||   :::   ||||||      :::
  394 spCysGly...ProLeuLysCysAspAsnHisCysCysAsnGlySerThr 409

 1594 TGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAA 1643
       |||:::|||   :::   |||:::|||:::::|||   ||||||:::
  410 CysLysLeuIleGlyGluAlaGluCysAlaSerGlyAspCysCysAspLe 426

 1644 TACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGA 1693
         :::|||      :::||||||:::      ||||||   |||::::
  426 uLysThrCysLysProLysProArgAlaThrValCysArgAlaAlaIleG 443

 1694 ACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCA 1743
       ::   ||||||:::   |||||||||:::|||:::::::::::::||||||
  443 lyIleCysAspLeuAspGluTyrCysAsnGlyGluThrAsnAspCysPro 459

 1744 CCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCG 1793
        ::::::      |||::::::      |||      ::::::::::
  460 AlaAspPhePheValGlnAsnAlaAlaLeuCysProGlyLysGluAsnGl 476
```

```
1794 C...TGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACA 1840
     :   ||||||:::|||   |||   :::|||:::::::||||||     :
 476 uPheCysTyrGluGlyGlyCysGlySerArgAsnAspGlnCysAlaLysL 493

1841 TCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTG 1890
     ::|||||||      :::   ::::::|||:::   ||||||:::|||
 493 euTrpGlyProThrGlyLysAsnGlyAspGluAsnCysTyrArgLys... 508

1891 AATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGA......GA 1934
     ||||||||||||||      ||||||||||||   ::::::
 509 AsnThrGluGlyThrPheHisGlyAsnCysGlyThrAsnAlaHisThrLy 525

1935 CCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCT 1984
       :::      :::|||:::   :::::::::   ||||||:::|||  |
 525 sGluIleLysLysCysGluThrGluAsnAlaLysCysGlyLeuLeuGlnC 542

 985 GTACCAATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATC 2034
     ||   :::   ::::::   |||   :::|||:::      |||:::::::
 542 ysGluThrGlnAlaGluArgProValPheGlyAsp...ProGlySerVal 557

2035 ATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGC 2084
             :::|||   |||                    |||   ::
 558 .........ThrPheSerHis.....................SerThrVa 564

2085 CCATGTAGTTTTAGATGATGAT.......................... 2106
     ::::      |||      |||
 564 lTyrSerSerLeuLysArgAspAspLysLysPheCysTyrValPheLysS 581

2107 ......................ACGGATGTGGGCTATGTAGAAGATGGA 2133
                            |||   |||   |||   ||||||
 581 erAlaTyrGlyGlyLeuAsnAlaProAspProGlyLeuValProAspGly 597

2134 ACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAAT 2183
     :::   ||||||   :::::|||||:::   :::||||||   :::
 598 AlaIleCysGlyGluGluGlnMetCysIleGlyGlnLysCysHisLysLy 614

2184 TCAAGCCCTAAAT......ATGAGCAGCTGTCCACTCGATTCCAAGGGTA 2227
      :::   ::::::         :::::::|||   ||||||:::
 614 sGluLysIleSerLysValThrAlaGlnCys...LeuAspAsn....... 627

2228 AAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGT 2277
          |||:::   :::|||||||||:::|||   ::::::|||   |||
 628 .....CysAsnPheArgGlyValCysAsnAsnValGlyAsnCysHisCys 642

2278 GATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGTTAGGAA 2327
     :::      ::::::|||      |||:::|||              ::
 643 GluArgGlyPheGlyGlyIleAlaCysGluIleProGlyTyrGlyGlySe 659

2328 CCTTCACCCCCCCAAGGATGAAGGACCCAAGGGT......CCTAGTGCCA 2371
     :::::::      :::          :::|||        |||:::|
 659 rValAsnSerAsnGluAlaTyrArgPheArgGlyIleThrLeuSerSerT 676

2372 CCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCTATT 2421
     ||   |||::::::      :::   |||                  |||
 676 hrPheLeuValPhePheCysLeuPheGly.................Ile 686

2422 GTCCTTGGGGGC..................................... 2433
        :::||||||
 687 PheIleGlyGlyLeuCysValTyrTyrArgValLysArgLysArgAsnIe 703

2434 ....ACAGGCTGGGGATTTAAAAATGTCAAGAAGAGAAGGTTCGAT.... 2475
         :::   |||         :::|||   |||::::::||||||
 703 uValSerGluTrpTrp......SerValValLysLysLysPheAspLeuH 718

2476 ..............CCTACTCAGCAAGGCCCC 2493
                   |||:::::::::::|||
 718 isGlyAspLeuValProValArgLysAlaPro 728
```

seq_name: pir2:S18968

seq_documentation_block:
cyritestin precursor - mouse
C;Species: Mus musculus (house mouse)

C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 21-Jan-2000
C;Accession: I48784; S18968
R;Senftleben, A.; Wallat, S.; Lemaire, L.; Heinlein, U.A.O.
Dev. Growth Differ. 36, 49-58, 1994
A;Title: Pre and postmeiotic germ cell specific expression of TAZ83, a gene encoding
A;Reference number: I48784
A;Accession: I48784
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-823 <RES>
A;Cross-references: EMBL:X64227; NID:g54264; PID:g54265
C;Genetics:
A;Gene: TAZ83
C;Superfamily: mouse meltrin alpha; disintegrin homology
F;395-480/Domain: disintegrin homology <DIS>

alignment_scores:
Quality: 658.50 Length: 693
Ratio: 1.719 Gaps: 27
Percent Similarity: 55.267 Percent Identity: 27.561

alignment_block:
US-09-634-252A-2 x S18968 ..

Align seg 1/1 to: S18968 from: 1 to: 823

```
 328 GAAATCACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGA 377
     :::||||||:::|||:::::::::      |||::::::   :::||
  28 GlnIleThrValProGluLysIleAspThrAsnIleGlnAspAlaLysGl 44

 378 AAGCCCTTATCACGTT..............CTTGACACAAAGGCAAGAC 412
     |:::      :::|||               ::::::   ||||||
  44 uAlaGluThrGlnValThrTyrValValThrIleGluGlyLysAla.... 59

 413 ACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAG 462
                          :::::::::|||   :::   ||||||
  60 ................TyrThrLeuGlnLeuGluLysGlnSerPheLeu 70

 463 ATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGG 512
           ||||||:::   ::::::   |||   :::         ||
  71 HisProLeuPheGlyThr...TyrLeuArgAspLysLeu.........Gl 83

 513 TTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAAATGGGAAACCAC 562
     |   |||                                    :::|||
  83 yThrLeu.....................................GlnProT 88

 563 AGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGC 612
       :::|||            ||||||:::|||:::|||
  88 yrPheSerLeuValLysThrHisCysPheTyrGlnGlyHisAlaAlaGlu 104

 613 GTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCAT 662
     :::      |||   |||:::|||||||||||||:::|||||:::|||::
 105 IleProValSerThrValThrLeuSerThrCysSerGlyLeuArgGlyLe 121

 663 GTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTC 712
     :::::::   ::::::::::::|||   ||||||||||||
 121 uLeuGlnLeuGluAsnIleThrTyrGlyIleGluProLeu.......... 134

 713 ATGATGAGAAAAGCACAGGTCGACCACATATAATCCAG......AAAACC 756
           |||:::|||:::       ||||||:::        |||:::
 135 .....GluSerSerAlaThrPheGluHisIleLeuTyrGluIleLysAsn 149

 757 TTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAGG 806
               :::||||||        :::|||
 150 AsnLysIleAspTyrSerProLeuLysGluAsn.................. 160

 807 TGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAAGA 856
                   |||   ::::::   |||      :::      :::
 161 ............PheAlaAsnSerGluGlnGluSerGlnSerTyrArgI 173

 857 GAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATG......AAATAT 900
           |||:::|||:::::::|||      :::   :::      :::
```

```
 173 leLeuValLysProGluLysGlySerAsnSerThrLeuThrLysArgIle 189

 901 TTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTC 950
     |||::::::   |||:::   |||         :::           ||
 190 LeuArgIleLysIleIleMetAspLysAlaMetPheAspHisMetGlySe 206

 951 TTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGG 1000
     |:::   :::   :::::::::   :::      :::   :::|||:::: 
 206 rGluValGlyValAlaThrGlnLysValValHisIlePheGlyLeuIleA 223

1001 ATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTA 1050
     ::::::::::   :::||||||:::      |||:::|||   ::::::
 223 snThrMetPhe...SerGlnLeuLysMetThrValMetLeuAsnSerLeu 238

1051 GAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCA 1100
     |||   |||:::|||:::|||:::|||:::   :::   :::      :: 
 239 GluIleTrpSerGluGlnAspLysIleGluThrAsnGlyAspAlaAspGl 255

1101 GATGCTCCATGAGTTCTCAAAATACCGGCAGCGC...ATTAAGCAGCATG 1147
     ::::|||::::::|||      ::::::::::   |||:::|||   |
 255 uValLeuGlnArgPheLeuLeuTrpLysSerLysGluIleSerGlnLysA 272

 148 CTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGC 1197
     ||:::   :::      :::               |||
 272 laGlnAspIleThrTyrLeuLeuLeuTyrLysAspHisProAspTyrVal 288

1198 AGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGTGT 1247
     :::   :::|||   |||   :::|||:::   :::      :::|||::
 289 GlyAlaThrTyrHisGlyMetAlaCysAsnProAsnPheThrAlaGlyIl 305

1248 GAATGAGTATGGTCTTCCAATGGCAGTG.........GCACAAGTATTAT 1288
     :      :::         :::|||||         |||   |||||||
 305 eAlaLeuHisProLysThrLeuAlaValGluGlyPheAlaIleValLeuS 322

1289 CGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGA 1338
     |||||   |||:::   |||||||||:::   ::::::
 322 erGlnLeuLeuGlyIleAsnLeuGlyLeuAlaTyrAspAspValTyr... 337

1339 AAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGA 1388
           :::|||   |||       |||         |||||||||:::
 338 ......AsnCysPheCysProGlySer......ThrCysIleMetAsnPr 350

1389 AACAGGGGTGTCCCATTCT......CGAAAATTTTCAAAGTGCAGCATTT 1432
      :::::::::      :::      :::   ||||||:::|||||:::
 350 oSerAlaIleArgSerGlnGlyIleLysValPheSerSerCysSerValA 367

1433 TGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAAC 1482
       |||:::::::::::::   ::::::            ||||||   |||
 367 spGluPheLysGlnLeuAlaSerGlnProGluLeuAspCysLeuArgAsn 383

1483 AGGCCAACAAAGCTATTT......GAGCCC.........ACGGAATGTGG 1517
                  |||       :::|||          :::   |||||
 384 ThrSerGluThrGluPheValGlnProGlnGlyGlySerTyrCysGl 400

1518 AAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGG 1567
     ||||      :::|||:::   |||:::::::::::::::
 400 yAsnHisLeuLeuGluValProGluGlnCysAspCysGlyProProGluT 417

1568 AATGC...TATGGATTATGCTGT......AAGAAATGTTCCCTCTCCAAC 1608
       |||   :::      ||||||      |||   |||:::|||   :::
 417 hrCysThrHisLysLysCysCysAsnProLysAspCysThrLeuIleAsp 433

1609 GGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTT 1658
     :::|||:::|||:::   ||||||||||||:::::::   :::|||   ::
 434 AlaAlaGlnCysGlyThrGlyProCysCysAspLysArgThrCysThrIl 450

1659 TCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTA 1708
     :      ||||||      ||||||   :::   ::::::|||||:::
 450 eAlaGluArgGlyArgLeuCysArgLysSerLysAspGlnCysAspPheP 467

1709 CTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAG 1758
       |||:::|||:::|||::::::      |||   |||:::
 467 roGluPheCysAsnGlyGluThrGluGlyCysAlaProAspThrLysAla 483

1759 CAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGA 1808
       |||         |||||::::::   :::   |||::::::|||
 484 AlaAspLeuGluProCysAsnAsnGluThrAlaTyrCysPheGlyGlyVa 500

1809 GTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTG 1858
       |||:::      |||:::||||||      ::::::|||      |||
 500 lCysArgAspProAspArgGlnCysThrAspLeuPheGlyLysTyrAlaL 517

1859 CAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAG 1908
       |||   :::      |||   :::::::::|||   ::::::   :::
 517 ysGlyProAsnTyrValCysAlaGlnGluValAsnLeuGlnAsnAspLys 533

1909 AAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACA 1958
       |||||||||                        :::|||:::
 534 PheGlyAsnCysHisGly.................ArgCysAsnTyrSe 544

1959 TGATGTGTTCTGTGGATTCTTACTCTGT...ACCAATCTTACTCGAGCTC 2005
         :::|||||||||      :::|||      |||:::
 544 rAlaIlePheCysGlyLysAlaValCysTyrTrpAsnPhe..... 557

2006 CACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCAT 2055
                           :::|||:::|||   |||:::   |||
 558 ...................AlaGluValIleGlnThrGluLysTyrAsp 567

2056 ..................CAAGGCCGGGTGATTGACTGCAGTGGTGCCCA 2087
                            |||:::|||      |||   :::|||||
 568 ValGlnTyrThrTyrLeuGlyGlyGlnVal......CysValSerAlaHi 582

2088 TGTAGTTTTAGATGATGATACG...GATGTGGGCTATGTAGAAGATGGAA 2134
     |:::      :::      |||   |||      ||||||:::|||||||
 582 sLeuArgSerGlnThrGlyThrArgAspAspThrTyrValHisAspGlyT 599

2135 CGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAATT 2184
     ||   ||||||   ::::::::::|||:::         ||||||::::::
 599 hrValCysGlySerGlyGlnValCysPheArgGlyAspCysLeuArgVal 615

2185 CAAGCCCTA...AATATGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGT 2231
     ::::::|||   :::      :::|||      |||:::|||
 616 HisValLeuArgGlyThrArgGluCysGluAlaAspAspLys........ 629

2232 CTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGATT 2281
       |||:::|||||||||:::|||:::|||      :::|||   |||:::
 630 .CysGlnGlyHisGlyIleCysAsnAsnLeuAsnAsnCysGlnCysGluS 646

2282 TCACCTGGGCAGGGACAGATTGCAGTATC 2310
           :::|||      :::|||::::::
 646 erGlyPheAlaProProGluCysAspMet 655
```

seq_name: pir2:I49281

seq_documentation_block:
fertilin alpha precursor - mouse (fragment)
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jun-2000
C;Accession: I49281
R;Wolfsberg, T.G.; Straight, P.D.; Gerena, R.L.; Huovila, A.P.; Primakoff, P.; Myles,
Dev. Biol. 169, 378-383, 1995
A;Title: ADAM, a widely distributed and developmentally regulated gene family encodin
A;Reference number: I48100; MUID:95269891
A;Accession: I49281
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-600 <RES>
A;Cross-references: EMBL:U22056; NID:g965009; PIDN:AAA74920.1; PID:g965010
C;Genetics:
A;Gene: ADAM 1
C;Superfamily: unassigned disintegrins; disintegrin homology
F;246-326/Domain: disintegrin homology <DIS>
F;180/Active site: Glu #status predicted

alignment_scores:

```
          Quality:  651.50            Length:    522
            Ratio:  2.068               Gaps:     20
Percent Similarity:  60.345   Percent Identity: 31.801

alignment_block:
US-09-634-252A-2 x I49281   ..

Align seg 1/1  to: I49281  from: 1  to: 600

     895 AAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCA 944
         ||||||:::|||:::::::::|||   :::|||:::   ::::::
          44 LysTyrValGluMetPheValValValAsnHisGlnArgPheGlnMetTr 60

     945 TCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACC 994
             |||:::      :::|||   :::   :::   :::::::::   |
          60 pGlySerAsnIleAsnGluThrValGlnAlaValMetAspIleIleAlaL 77

     995 TTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTG 1044
         ||:::::|||   :::   :::   :::||||||:::||||||||||||
          77 euAlaAsnSer...PheThrArgGlyIleAsnThrGluValValLeuVal 92

    1045 GCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCC 1094
         ::::::|||   |||||||||  |||   |||:::::   ::::::
          93 GlyLeuGluIleTrpThrGluGlyAspProIleGluValProValAspLe 109

    1095 TGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAG..... 1131
                 |||:::::::|||:::   :::||||||
         109 uGlnThrThrLeuArgAsnPheAsnPheTrpArgGlnGluLysLeuValG 126

    1132 ..CGCATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACA 1179
           |||::::::::::      |||:::::::|||||||||          :::
         126 lyArgValArgHis......AspValAlaHisLeuIle.........Val 137

    1180 TTTCACTAT......AAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTG 1223
           |||          ::::::::::::   :::::::::   |||:::||
         138 GlyHisArgProGlyGluAsnGluGlyGlnAlaPheLeuArgGlyAlaCy 154

    1224 TTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCA......A 1267
         ||||          ::::::::::|||:::   :::                :
         154 sSerGlyGluPheAlaAlaAlaValGluAlaPheHisHisGluAspValL 171

    1268 TGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATC 1317
         ::      |||   ::::::::::::::|||:::::|||||||||||
         171 euLeuPheAlaAlaLeuMetAlaHisGluLeuGlyHisAsnLeuGlyIle 187

    1318 CAATGGGAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTG 1367
         |||   :::                |||   |||   |||      :::
         188 GlnHisAsp...........HisProThrCysThrCysGlyProLys.. 199

    1368 GGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAATTTT 1417
                |||:::|||   |||      ::::::   :::      ||||
         200 .HisPheCysLeuMetGlyGluLys...IleGlyLysAspSerGlyPheS 215

    1418 CAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGA 1467
         ||:::||||||      ::::::      ||||||:::          |||
         215 erAsnCysSerSerAspHisPheLeuArgPheLeuHisAspHisArgGly 231

    1468 GCCTGCCTTTTCAACAGGCCA.........ACAAAGCTATTTGAGCCCAC 1508
         |||||||||:::::::::|||          ::::::::   :::   ::
         232 AlaCysLeuLeuAspGluProGlyArgGlnSerArgMetArgArgAlaAl 248

    1509 GGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTT 1558
         ::::|||||||||||||   ||||||      ||||||||||||||||||
         248 aAsnCysGlyAsnGlyValValGluAspLeuGluGluCysAspCysGlyS 265

    1559 TTCATGTGGAATGCTATGGATTATGCTGTAAGAAATGTTCCCTCTCCAAC 1608
                 :::   :::      |||          |||:::|||::::::
         265 erAspCysAspSerHisProCysCysSerProThrCysThrLeuLysGlu 281

    1609 GGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTT 1658
         ||||||:::||||||:::|||   ||||||   |||      |||   ||
         282 GlyAlaGlnCysSerGluGlyLeuCysCysTyrAsn......CysThrPh 296

    1659 TCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTA 1708
         |:::   :::|||      ||||||   |||   :::   ||||||:::
          296 eLysLysLysGlySerLeuCysArgProAlaGluAspValCysAspLeuP 313

    1709 CTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAG 1758
           |||||||||   |||::::::   :::||||||   |||
          313 roGluTyrCysAspGlySerThrGlnGluCysProAlaAsnSerIleMet 329

    1759 CAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGA 1808
         |||||||||      |||::::::            |||   :::|||
          330 GlnAspGlyThrGlnCysAspArgIleTyr...TyrCysLeuGlyGlyTr 345

    1809 GTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTG 1858
          ||||||:::   |||:::|||||||:::   |||:::|||      |||
          345 pCysLysAsnProAspLysGlnCysSerArgIleTyrGlyTyrProAlaA 362

    1859 CAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAG 1908
           ::::::   :::   ||||||   ::::::||||||:::::::::::::
          362 rgSerAlaProGluGluCysTyrIleSerValAsnThrLysAlaAsnArg 378

    1909 AAGGGAAACTGCGGGAAGGATGGAGAC.........CGGTGGATTCAGTG 1949
           ||||||||||||          :::         |||:::      ||
          379 PheGlyAsnCysGlyHisProThrSerAlaAsnPheArgTyrGluThrCy 395

    1950 CAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTC 1999
         ||||   :::|||||||||||||||   |||:::||||||::::::
          395 sSerAspGluAspValPheCysGlyLysLeuValCysThrAspValArgT 412

    2000 GAGCTCCACGTATTGGTCAACTTCAGGGTGAGATC...ATTCCAACTTCC 2046
              |||::::::      |||::::::   :::   :::|||
          412 yrLeuProLysValLysProLeuHisSerLeuLeuGlnValPro..... 426

    2047 TTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTT 2096
             |||   :::            :::   ::::::   :::::::::
          427 ...TyrGlyGluAspTrpCysTrpSerMetAspAlaTyrAsnIleThr.. 441

    2097 AGATGATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCC 2146
             |||      |||   |||   |||::::::|||:::   |||:::|
          442 ....AspValProAspAspGlyAspValGlnSerGlySerPheCysAlaP 457

    2147 CGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAAT 2196
         ||:::   :::|||::::::      |||            :::|||:::
          457 roAsnLysValCysMetGluTyrIleCysThrGlyArgGlyValLeuGln 473

    2197 ATGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGG 2246
           :::      |||   :::            :::|||   |||:::||
          474 TyrAsnCysGluProGlnGlu.............MetCysHisGlyAsnGl 486

    2247 GGTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGA 2296
         |||||||:::|||          |||   ||||||      :::|||
          486 yValCysAsnAsnPheLysHisCysHisCysAspAlaGlyPheAlaProP 503

    2297 CAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAAG... 2343
           |||||||||                                |||   :::
          503 roAspCysSer.....................SerProGlyAsnGly 511

    2344 .........GATGAAGGACCCAAGGGT...CCTAGTGCCACCAATCTCAT 2381
                   |||:::||||||   |||   |||:::      :::|||
          512 GlySerValAspSerGlyProValGlyLysProAlaAspArgHisLeuSe 528

    2382 AATAGGCTCCATCGCT 2397
          ::::::   :::|||
          528 rLeuSerPheLeuAla 533

seq_name: pir2:S48169

seq_documentation_block:
metalloproteinase (EC 3.4.24.-) H-II precursor - carpet viper (fragment)
N;Contains: disintegrin
C;Species: Echis pyramidum leakeyi
C;Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 09-Jun-2000
C;Accession: S48169
R;Paine, M.J.I.; Moura-Da-Silva, A.M.; Theakston, R.D.G.; Crampton, J.M.
```

```
Eur. J. Biochem. 224, 483-488, 1994
A;Title: Cloning of metalloprotease genes in the carpet viper (Echis pyramidum leakeyi).
A;Reference number: S48160; MUID:95010025
A;Accession: S48169
A;Molecule type: mRNA
A;Residues: 1-549 <PAI>
A;Cross-references: EMBL:X78971; NID:g763094; PIDN:CAA55566.1; PID:g763095
C;Superfamily: mouse meltrin alpha; disintegrin homology
C;Keywords: hydrolase; metalloproteinase; venom; zinc
F;1-138/Domain: propeptide (fragment) #status predicted <PRO>
F;139-549/Product: metalloproteinase H-II #status predicted <MAT>
F;350-432/Domain: disintegrin homology <DIS>
F;284/Active site: Glu #status predicted


alignment_scores:
            Quality:  647.00              Length:    605
              Ratio:  1.875                 Gaps:     25
 Percent Similarity:  57.025   Percent Identity: 30.248

[redacted]gment_block:
[redacted]9-634-252A-2 x S48169   ..

 Align seg 1/1  to: S48169  from: 1  to: 549

    457 TTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAA 506
        ::::::::::::::::   |||:::      :::|||   |||      ||
      5 TyrGluPheLysValLysGlyGluProValValLeuHisLeuGlnLysAs 21

    507 CAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAAATGGGA 556
        |:::   |||:::|||:::||||||   |||   ||||||
     21 nLysAspLeuPheSerGluAspTyrSerGluThrHisTyrThrProAspG 38

    557 AACCACAGTACTCTAAGGGT.........GGAGAGCACTGTTACTACCAT 597
            :::   :::   :::            :::|||||||||||||||
     38 lyArgGluIleThrThrAsnProProValGluAspHisCysTyrTyrHis 54

    598 GGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAA 647
        |||   :::::::::          |||:::::::::::::|||:::|||||
     55 GlyArgValGlnAsnAspAlaHisSerSerAlaSerIleSerAlaCysAs 71

    648 TGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGC 697
        |||||||   |||   |||:::   :::       :::|||:::|||||||
     71 nGlyLeuLysGlyHisPheLysLeuGlnGlyGluThrTyrPheIleGluP 88

    698 CACTAGAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCACATATAATC 747
        |||||:::::::       :::|||                |||   :::
     88 roLeuLysIleProAspSerGlu..............AlaHisAlaVal 99

    748 CAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTAT 797
           |||                 |||:::                 :::::
    100 TyrLys..............TyrGlu.................AsnIl 105

    798 GGAAAGAGGTGACCAGTGGCCCTTTCTC.........TCTGAATTACAGT 838
        :|||:::   ||||||   |||   :::         ::::::   :::|
    105 eGluLysGluAspGlnAlaProLysMetCysGlyValThrHisThrAsnT 122

    839 GG..............TTGAAAAGAAGGAAGAGAGCAGTGAATCCATCA 873
        ||               :::|||:::   :::|||   |||      |||
    122 rpGluSerAspGluProIleLysGluAlaSerArgLeuValAlaSerSer 138

    874 ......CGTGGTATATTTGAAGAAATGAAATATTTGGAACTTATGATTGT 917
              ::::::   :::::::::::::::::|||::::::|||:::|||||
    139 GluGlnGlnSerTyrTyrAspAsnPheArgTyrValLysLeuValIleVa 155

    918 TAATGATCACAAAACGTATAAGAAGCATCGCTCT...TCTCATGCACATA 964
        |   ||||||:::         |||:::::::::   |||      :::|
    155 lValAspHisArgMetValThrLysTyrAsnAsnAspSerThrIleArgT 172

    965 CCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAG 1014
        ||:::   :::      ::::::||||||   |||::::::||||||
    172 hrArgIleTyr......GluMetValAsnThrValAsnGluIleTyr... 185

   1015 GAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGA 1064
           :::||||||   ||||||:::||||||:::::|||   |||:::::
    186 IleHisLeuAsnIleArgValAlaLeuValGlyLeuGluPheTrpSerAs 202

   1065 GAAGGATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGT 1114
        :   |||   |||::::::|||:::         :::   |||:::   |
    202 nGlyAspLeuIleAsnValThrSerAlaAlaGluHisThrLeuAsnLeuP 219

   1115 TCTCAAAATACCGGCAGCGC.........ATTAAGCAGCATGCTGATGCT 1155
        ||:::   :::|||                    :::::::::   |||
    219 heGlyValTrpArgAlaSerAspLeuLeuSerArgArgArgHisAspAsn 235

   1156 GTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAG 1205
        :::||||||::::::   :::   ::::::             :::|||
    236 AlaHisLeuLeuThrAlaIleAspLeuAsnGlyProThrIleGlyLeuAr 252

   1206 TTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGT 1255
              :::::::::|||:::   ||||||:::|||||||||   :::
    252 gAspValSerSerMetCysGlnAlaThrArgSerValGlyValValGlnA 269

   1256 ATGGTCTTCCA......ATGGCAGTGGCACAAGTATTATCGCAGAGCCTG 1299
              |||         |||||||||   ::::::::::::::::::
    269 spHisSerProThrValArgAlaValAlaValThrMetAlaHisGluMet 285

   1300 GCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATG 1349
        :::::|||||||||:::::   :::   :::                ||
    286 GlyHisAsnLeuGlyMetSerHisAspGlyAsn............HisCy 298

   1350 TGACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGT 1399
        |:::|||      :::      :::|||||||||      :::      :
    298 sAsnCysGlyAlaAsn......SerCysIleMetAlaAlaValLeuArgA 313

   1400 CCCATTCTCGAAAA...TTTTCAAAGTGCAGCATTTTGGAGTATAGAGAC 1446
        ::   :::   :::   ||||||   ||||||       |||::::::
    313 snProAlaProGluTyrPheSerAspCysSerArgArgTyrTyrGlnAsn 329

   1447 TTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAG.. 1494
        ||||||   :::            |||   :::   ||||||:::|||
    330 PheLeuThrAsnTyrIleProAspCysThrIleIleArgProSerLysTh 346

   1495 ....CTATTTGAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGG 1540
           :::   :::|||      ||||||||||||   ::::::   ||||
    346 rAspIleValSerProGlnValCysGlyAsnGlyLeuLeuAspGluGlyG 363

   1541 AGGAGTGTGATTGTGGTTTTCATGTGGAATGCTATGGATTATGCTGT... 1587
        ||||||||||||||||      :::::::|||          ||||||
    363 luGluCysAspCysGlySerProAlaAsnCysGlnTyrProCysCysAsp 379

   1588 ...AAGAAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTG 1634
           :::|||:::|||   :::   ::::::|||:::   |||   ||
    380 AlaAlaSerCysLysLeuHisSerTrpValGluCysGluPheGlyHisCy 396

   1635 CTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGG 1684
        ||||:::      :::|||   |||:::|||   |||   |||||||||
    396 sCysAsp......GlnCysArgPheLysProAlaGlyThrGluCysArgG 411

   1685 ATGCTGTGAACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGT 1734
              :::||||||||||:::   ||||||||||||||||:::|||:::
    411 lyIleArgSerGluCysAspLeuProGluTyrCysThrGlyGlnSerAla 427

   1735 CAGTGCCCACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAA 1784
        :::|||    ::::::|||   ::: |||         |||      |||
    428 GluCys.LeuArgMetPheSerThr.ArgTrpLysThrMet...ProLys 442

   1785 TCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTC 1834
           :::   ||||||||||||    |||       :::   ||||||
    443 LeuSerLeuLeuLeuGlnTrp.ThrCysProIleMetGlnTyrGlnCysT 459

   1835 AGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAA 1884
              :::|||   :::|||:::   ::::::      |||:::|||
    459 yrAlaHisPheGlyGlnAsnAlaValValGlyGlnAspAlaCysPheGlu 475

   1885 AAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGA 1934
```

```
          :::|||   ||||||               |||   |||:::::::||
    476 ...IleAsnLysGluGlyLysGlyAspPheTyrCysArgLysGluAsnAs 491

   1935 CCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCT 1984
        |     |||   |||:::::::::||||||   ||||||   |||:::|
    491 pValProIleProCysAlaGlnGluAspValLysCysGlyArgLeuPheC 508

   1985 GT.........ACCAATCTTACTCGAGCTCCACGTATTGGTCAACTTCAG 2025
        ||           |||:::   |||   |||
    508 ysGluThrGluProAsnMetCysArgTyrProTyr..............  519

   2026 GGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTGACTG 2075

    519 ..................................................  519

    076 CAGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTATGTAG 2125
                                        |||   |||   |||:
    520 ..............................GlyAspGluGlyMetValA 526

   2126 AAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGC 2175
        .  ::   ||||||   |||      :::   :::|||:::::::   ||||||
    526 spProGlyThrLysCysGluAspLysLysValCysIleAsnGlyLysCys 542

   2176 CTACAAATT 2184
        :::::::::
    543 IleAspVal 545
```

seq_name: pir2:A37877

seq_documentation_block:
hemorrhagic protein HR1B - habu
C;Species: Trimeresurus flavoviridis (habu)
C;Date: 31-Jul-1991 #sequence_revision 31-Jul-1991 #text_change 19-May-2000
C;Accession: A37877
R;Takeya, H.; Oda, K.; Miyata, T.; Omori-Satoh, T.; Iwanaga, S.
J. Biol. Chem. 265, 16068-16073, 1990
A;Title: The complete amino acid sequence of the high molecular mass hemorrhagic protein
A;Reference number: A37877; MUID:90375459
A;Accession: A37877
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-416 <TAK>
C;Superfamily: mouse meltrin alpha; disintegrin homology
F;210-292/Domain: disintegrin homology <DIS>
F;144/Active site: Glu #status predicted

[redacted]nment_scores:
Quality: 610.00 Length: 443
Ratio: 2.302 Gaps: 16
Percent Similarity: 59.819 Percent Identity: 31.828

alignment_block:
US-09-634-252A-2 x A37877 ..

Align seg 1/1 to: A37877 from: 1 to: 416

```
    895 AAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCA 944
        :::|||:::::|||   ||||||   ||||||            |||||
      7 ArgTyrIleLysLeuAlaIleValValAspHisGlyIleValThrLysHi 23

    945 TCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACC 994
        |:::::::::           :::::::::      ::::::||||||
     23 sHisGlyAsnLeuLysLysIleArgLysTrpIleTyrGlnLeuValAsnT 40

    995 TTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTG 1044
          :::::::::||||||:::   :::||||||      |||:::||||||
     40 hrIleAsnAsnIleTyrArg...SerLeuAsnIleLeuValAlaLeuVal 55

   1045 GCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCC 1094
          :::|||   |||:::::::::::::::|||   :::   :::
     56 TyrLeuGluIleTrpSerLysGlnAsnLysIleThrValGlnSerAlaSe 72

   1095 TGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATT......A 1138
                  |||      |||:::   :::|||:::   :::      |
     72 rAsnValThrLeuAspLeuPheGlyAspTrpArgGluSerValLeuLeuL 89

   1139 AGCAGCATGCT...GATGCTGTGCACCTCATCTCGCGGGTGACATTTCAC 1185
        |||||::::::   |||::::::::::|||::::::   :::   |||
     89 ysGlnArgSerHisAspCysAlaGlnLeuLeuThrThrIleAspPheAsp 105

   1186 TATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAG 1235
                  :::   :::|||   :::::::::|||:::      ||
     106 GlyProThrIleGlyLysAlaTyrThrAlaSerMetCysAspProLysAr 122

   1236 AGGAGTTGGT...GTGAATGAGTATGGT...CTTCCAATGGCAGTGGCAC 1279
        |:::||||||   |||:::::|||:::   :::   ::::::||||||
     122 gSerValGlyIleValGlnAspTyrSerProIleAsnLeuValValAlaV 139

   1280 AAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCT 1329
          :::::::::::::::::::::::::||||||||||||      :::
     139 alIleMetThrHisGluMetGlyHisAsnLeuGlyIleProHisAspGly 155

   1330 TCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGC..... 1374
        :::|||            |||   |||           ||||||
     156 AsnSer............CysThrCys............GlyGlyPhePr 164

   1375 .TGCATCATGGAGGAAACAGGGGTGTCCCAT......TCTCGAAAATTTT 1417
         |||||||||::         :::|||         |||:::   ||||
     164 oCysIleMetSerProMet...IleSerAspProProSerGluLeuPheS 180

   1418 CAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGA 1467
        ||:::||||||         |||:::   ||||||
     180 erAsnCysSerLysAlaTyrTyrGlnThrPheLeuThrAspHisLysPro 196

   1468 GCCTGCCTTTTCAACAGGCCAACAAAG......CTATTTGAGCCCACGGA 1511
          |||:::::::|||   |||:::|||      :::   :::|||
     197 GlnCysIleLeuAsnAlaProSerLysThrAspIleValSerProProVa 213

   1512 ATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTC 1561
         |||||||||      :::||||||||||||||||||||||||||||
     213 lCysGlyAsnGluLeuLeuGluAlaGlyGluGluCysAspCysGlySerP 230

   1562 ATGTGGAATGCTATGGATTATGCTGT......AAGAAATGTTCCCTCTCC 1605
             :::|||          ||||||          :::|||:::|||
     230 roGluAsnCysGlnTyrGlnCysCysAspAlaAlaSerCysLysLeuHis 246

   1606 AACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCT 1655
        :::   :::   |||:::::::|||   |||||||::      :::|||
     247 SerTrpValLysCysGluSerGlyGluCysCysAsp......GlnCysAr 261

   1656 TTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATA 1705
         |||:::      |||   |||||||||   |||   :::||||||||||
     261 gPheArgThrAlaGlyThrGluCysArgAlaAlaGluSerGluCysAspI 278

   1706 TTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCAT 1755
        ||   |||   |||||||||:::|||::::::||||||   :::
     278 leProGluSerCysThrGlyGlnSerAlaAspCysProThrAspArgPhe 294

   1756 AAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGG 1805
          :::::::|||      |||      |||:::|||   ||||||||||
     295 HisArgAsnGlyGlnProCysLeuTyrAsnHisGlyTyrCysTyrAsnGl 311

   1806 CGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGG 1855
        |:::|||               ||||||   ::::::::|||:::::::|
     311 yLysCysProIleMetPheTyrGlnCysTyrPheLeuPheGlySerAsnA 328

   1856 CTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACT 1905
        ||:::   ::::::      |||:::::::::   |||   :::|||
     328 laThrValAlaGluAspAspCysPheAsnAsn...AsnLysLysGlyAsp 343

   1906 GAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAA 1955
        :::        |||   |||:::::::::::::::::|||   |||:::::
     344 LysTyrPheTyrCysArgLysGluAsnGluLysTyrIleProCysAlaGl 360

   1956 ACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTC 2005
```

```
     ::::||||||   ||||||   |||:::|||   |||
  360 nGluAspValLysCysGlyArgLeuPheCysAspAsn..... 372

 2006 CACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCAT 2055

  372 ..................................................  372

 2056 CAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGA 2105
         :::         |||         |||         ::::::||
  373 ......LysLysTyrProCys.........HisTyrAsnTyrSerGluAs 384

 2106 TACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGA 2155
      |   |||   |||   |||:::   ||||||   |||:::   :::   :
  384 pLeuAspPheGlyMetValAspHisGlyThrLysCysAlaAspGlyLysV 401

 2156 TGTGTTTAGATCGGAAGTGCCTACAAATT 2184
      ::|||   :::|||:::|||:::::::::
  401 alCysSerAsnArgGlnCysValAspVal 410

[redacted]name: pir2:A42972

[redacted]ocumentation_block:
[redacted]ulation factor X activating enzyme (EC 3.4.24.-) heavy chain - Russell's viper
R;Alternate names: RVV-X-heavy chain
C;Species: Vipera russelli (Russell's viper)
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 19-May-2000
C;Accession: A42972
R;Takeya, H.; Nishida, S.; Miyata, T.; Kawada, S.; Saisaka, Y.; Morita, T.; Iwanaga, S.
J. Biol. Chem. 267, 14109-14117, 1992
A;Title: Coagulation factor X activating enzyme from Russell's viper venom (RVV-X). A nc
A;Reference number: A42972; MUID:92332516
A;Contents: V. r. siamensis
A;Accession: A42972
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-429 <TAK>
A;Experimental source: venom
A;Note: sequence extracted from NCBI backbone (NCBIP:108405)
C;Comment: This metalloproteinase has disintegrin (platelet aggregation inhibitor)-like
C;Superfamily: mouse meltrin alpha; disintegrin homology
C;Keywords: hydrolase; metalloproteinase; venom; zinc
F;212-296/Domain: disintegrin homology <DIS>
F;146/Active site: Glu #status predicted

alignment_scores:
           Quality:  595.00            Length:    450
             Ratio:  2.095               Gaps:     19
 Percent Similarity:  63.111   Percent Identity: 31.778

alignment_block:
[redacted]9-634-252A-2 x A42972   ..

[redacted]n seg 1/1  to: A42972  from: 1  to: 429

  898 TATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCG 947
      ::::::||||||:::|||:::   ||||||:::      ||||||   ::
   12 PheIleGluLeuValIleIleValAspHisSerMetAlaLysLysCysAs 28

  948 CTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTG 997
      :|||:::   ::::::|||:::   :::      ::::::||||||   :
   28 nSerThrAlaThrAsnThrLysIleTyr......GluIleValAsnSerA 43

  998 TGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCT 1047
      :::::::::|||:::   :::   ||||||   :::|||:::|||:::::
   43 laAsnGluIlePhe...AsnProLeuAsnIleHisValThrLeuIleGly 58

 1048 GTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGT 1097
      ||||||   |||   :::::::|||   |||:::::::|||::::::
   59 ValGluPheTrpCysAspArgAspLeuIleAsnValThrSerSerAlaAs 75

 1098 GCAGATGCTCCATGAGTTCTCAAAATACCGG.....CAGCGCA 1135
       :::   |||::::::|||::::::::::|||      :::
   75 pGluThrLeuAsnSerPheGlyGluTrpArgAlaSerAspLeuMetThrA 92

 1136 TTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCAC 1185
        |||:::|||   :::|||:::   |||::::::   :::   |||
   92 rgLysSerHisAspAsnAlaLeu...LeuPheThrAspMetArgPheAsp 107

 1186 TATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAG 1235
        :::      ::::::::::::::::::|||:::|||:::      ||
   108 LeuAsnThrLeuGlyIleThrPheLeuAlaGlyMetCysGlnAlaTyrAr 124

 1236 AGGAGTTGGT...GTGAATGAGTATGGT...CTTCCAATGGCAGTGGCAC 1279
      |:::|||      |||:::|||   |||         :::   :::|||
   124 gSerValGluIleValGlnGluGlnGlyAsnArgAsnPheLysThrAlaV 141

 1280 AAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCT 1329
        ::::::::::::::|||::::::|||||||||:::      :::
   141 alIleMetAlaHisGluLeuSerHisAsnLeuGlyMetTyrHisAspGly 157

 1330 TCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCAT 1379
      ::::::            |||   |||::::::|||      :::|||::
   158 LysAsn............CysIleCysAsnAspSer......SerCysVa 168

 1380 CATGGAGGAAACA...GGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCA 1426
      :|||:::   :::   :::   :::   |||:::   ||||||:::||||
   168 lMetSerProValLeuSerAspGlnProSerLysLeuPheSerAsnCysS 185

 1427 GCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTT 1476
      |||||   :::|||:::   :::|||   |||            |||:::
   185 erIleHisAspTyrGlnArgTyrLeuThrArgTyrLysProLysCysIle 201

 1477 TTCAACAGGCCAACAAAG......CTATTTGAGCCCACGGAATGTGGAAA 1520
      ||||||   |||   :::      :::   :::|||      ||||||||
   202 PheAsnProProLeuArgLysAspIleValSerProProValCysGlyAs 218

 1521 TGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAAT 1570
      |         |||   |||||||||||||||||||||      ::::::|
   218 nGluIleTrpGluGluGlyGluGluCysAspCysGlySerProAlaAsnC 235

 1571 GCTATGGATTATGCTGT......AAGAAATGTTCCCTCTCCAACGGGGCT 1614
      ||   :::   |||:||            |||:::|||:::   ||||||
   235 ysGlnAsnProCysCysAspAlaAlaThrCysLysLeuLysProGlyAla 251

 1615 CACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCC 1664
      :::|||:::::::|||   ||||||         :::|||   ::::::
   252 GluCysGlyAsnGlyLeuCysCys......TyrGlnCysLysIleLysTh 266

 1665 ACGAGGGTATGAATGCCGGGAT......GCTGTGAACGAGTGTGATATTA 1708
         |||      ||||||         |||   :::||||||||:::
   266 rAlaGlyThrValCysArgThrArgArgAlaArgAspGluCysAspValP 283

 1709 CTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAG 1758
        |||:::||||||||:::|||::::::||||||   :::      :::
   283 roGluHisCysThrGlyGlnSerAlaGluCysProArgAspGlnLeuGln 299

 1759 CAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGA 1808
      |||:::|||      |||:::::|||:::|||   ||||||||||||::
   300 GlnAsnGlyLysProCysGlnAsnAsnArgGlyTyrCysTyrAsnGlyAs 316

 1809 GTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTG 1858
      :|||            |||||||||      :::::::|||:::::|||
   316 pCysProIleMetArgAsnGlnCysIleSerLeuPheGlySerArgAlaA 333

 1859 CAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAG 1908
         :::         |||::::::::::   |||   :::|||:::
   333 snValAlaLysAspSerCysPheGlnGlu...AsnLeuLysGlySerTyr 348

 1909 AAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACA 1958
        |||   |||   |||::::::   |||   |||   |||:::   ::
   349 TyrGlyTyrCysArgLysGluAsnGlyArgLysIleProCysAlaProGl 365

 1959 TGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCAC 2008
      :||||||   ||||||   |||:::|||   |||      ::::::||||
   365 nAspValLysCysGlyArgLeuPheCysLeuAsn......AsnSerProA 380
```

```
 2009 GTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAA 2058
      ||      :::                  |||   ::::::::::::
  380 rgAsnLysAsn.................ProCysAsnMetHisTyr... 389

 2059 GGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATAC 2108
                ::::||| ||
  390 ...........SerCysMetAspGlnHis..................... 395

 2109 GGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGT 2158
           |||  |||:::  ||||||  |||    :::  :::|
  396 ....LysGlyMetValAspProGlyThrLysCysGluAspGlyLysValC 411

 2159 GTTTAGAT...CGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCA3C 2205
      ||   :::   |||:::|||:::::::::::::::      :::|||:::
  411 ysAsnAsnLysArgGlnCysValAspValAsnThrAlaTyrGlnSerThr 427

 ame: pir2:S41607

seq_documentation_block:
atrolysin A (EC 3.4.24.1) - western diamondback rattlesnake (fragment)
N;Alternate names: hemorrhagic toxin a
C;Species: Crotalus atrox (western diamondback rattlesnake)
C;Date: 29-Sep-1994 #sequence_revision 13-Mar-1997 #text_change 09-Jun-2000
C;Accession: S41607
R;Hite, L.A.; Jia, L.G.; Bjarnason, J.B.; Fox, J.W.
Arch. Biochem. Biophys. 308, 182-191, 1994
A;Title: cDNA sequences for four snake venom metalloproteinases: structure, classificati
A;Reference number: S41607; MUID:94145078
A;Accession: S41607
A;Status: preliminary; translation not shown
A;Molecule type: mRNA
A;Residues: 1-419 <HIT>
A;Cross-references: EMBL:U01234; NID:g402257; PID:g402258
C;Superfamily: mouse meltrin alpha; disintegrin homology
C;Keywords: hydrolase; metalloproteinase
F;209-291/Domain: disintegrin homology <DIS>
F;143/Active site: Glu #status predicted


alignment_scores:
            Quality:  577.50              Length:    443
              Ratio:  2.213                Gaps:     14
 Percent Similarity:  58.916   Percent Identity: 31.151

alignment_block:
[redacted]-09-634-252A-2 x S41607   ..

[redacted]gn seg 1/1  to: S41607  from: 1  to: 419

  895 AAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCA 944
      :::|||:::||||||:::||||||  ||||||:::  :::   |||::
    6 ArgTyrValGluLeuValIleValAlaAspHisArgMetPheThrLysTy 22

  945 TCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACC 994
      ::::::::::           :::::::::     ::::::||||||
   22 rAsnGlyAsnLeuLysLysIleArgLysTrpIleTyrGlnIleValAsnT 39

  995 TTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTG 1044
        :::::::::||||||         ||||||  ||||||:::||||||
   39 hrIleAsnGluIleTyr...IleProLeuAsnIleArgValAlaLeuVal 54

 1045 GCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCC 1094
        :::|||   |||::::::   |||   ||||||:::|||:::
   55 ArgLeuGluIleTrpSerAsnGlyAspLeuIleAspValThrSerAlaAl 71

 1095 TGTGCAGATGCTCCATGAGTTCTCAAAATACCGG.............CAGC 1132
                |||   :::|||:::::::::|||              |
   71 aAsnValThrLeuLysSerPheGlyAsnTrpArgValThrAsnLeuLeuA 88

 1133 GCATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTT 1182
      ||   |||:::|||   |||   :::::::|||::::::   :::   :::
   88 rgArgLysSerHis...AspAsnAlaGlnLeuLeuThrAlaIleAspLeu 103

 1183 CACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCAC 1232
            :::      :::|||:::   :::|||   :::|||:::
  104 AspGluGluThrLeuGlyLeuAlaProLeuGlyThrMetCysAspProLy 120

 1233 AAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCA......ATGGCAGTGG 1276
      ::::::|||:::  :::          |||      :::   ||||
  120 sLeuSerIleGlyIleValGlnAspHisSerProIleAsnLeuLeuValA 137

 1277 CACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAA 1326
      ||   :::::::::::::::|||::::::|||||||||:::      :::
  137 laValThrMetAlaHisGluLeuGlyHisAsnLeuGlyMetValHisAsp 153

 1327 CCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTG 1376
         :::             :::|||   |||      |||      :::||
  154 GluAsn............ArgCysHisCys......SerThrProAlaCy 164

 1377 CATCATGGAGGAAACAGGGGTGTCCCATTCTCGA.........AAATTTT 1417
      |:::|||         :::|||   ::::::|||         :::||||
  164 sValMet......CysAlaValLeuArgGlnArgProSerTyrGluPheS 179

 1418 CAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGA 1467
      ||   ||||||:::   :::||||||   |||:::   :::   :::
  179 erAspCysSerLeuAsnHisTyrArgThrPheIleIleAsnTyrAsnPro 195

 1468 GCCTGCCTTTTCAACAGGCCA......ACAAAGCTATTTGAGCCCACGGA 1511
         |||::::::|||:::|||      |||   :::::::::|||
  196 GlnCysIleLeuAsnGluProLeuGlnThrAspIleIleSerProProVa 212

 1512 ATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTC 1561
       |||||||||      :::|||:::|||||||||||||||||||||
  212 lCysGlyAsnGluLeuGluLeuGlyGluGluCysAspCysGlySerP 229

 1562 ATGTGGAATGCTATGGATTATGCTGT......AAGAAATGTTCCCTCTCC 1605
            |||         ||||||          |||:::|||
  229 roArgThrCysArgAspProCysCysAspAlaAlaThrCysLysLeuHis 245

 1606 AACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCT 1655
      :::   ::::::|||:::::::|||   ||||||:::      :::|||
  246 SerTrpValGluCysGluSerGlyGluCysCysGln......GlnCysLy 260

 1656 TTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATA 1705
       |||         |||      ||||||   |||   :::||||||||||
  260 sPheThrSerAlaGlyAsnValCysArgProAlaArgSerGluCysAspI 277

 1706 TTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCAT 1755
      ||:::|||   |||||||||:::|||::::::||||||   :::
  277 leAlaGluSerCysThrGlyGlnSerAlaAspCysProThrAspAspPhe 293

 1756 AAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGG 1805
         :::::::|||      |||   :::|||   |||   ||||||||||
  294 HisArgAsnGlyLysProCysLeuHisAsnPheGlyTyrCysTyrAsnGl 310

 1806 CGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGG 1855
      |:::|||             :::||||||      :::||||||:::::::
  310 yAsnCysProIleMetTyrHisGlnCysTyrAlaLeuTrpGlySerAsnV 327

 1856 CTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACT 1905
      :::::   :::          |||::::::   :::|||   :::|||:::
  327 alThrValAlaProAspAlaCysPheAsp...IleAsnGlnSerGlyAsn 342

 1906 GAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAA 1955
      ::::::       |||   |||::::::          |||   |||:::::
  343 AsnSerPheTyrCysArgLysGluAsnGlyValAsnIleProCysAlaGl 359

 1956 ACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTC 2005
      ::::||||||   ||||||   |||:::|||
  359 nGluAspValLysCysGlyArgLeuPheCys..................... 369

 2006 CACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCAT 2055

  369 ..................................................... 369
```

```
    2056 CAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGA 2105
                ::::|||   |||          |||         ||||||
     370 ......AsnValAsnAspPheLeuCysArgHisLysTyrSerAspAsp.. 383

    2106 TACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGA 2155
                   |||   |||:::   ||||||   |||:::   :::   :
     384 ..........GlyMetValAspHisGlyThrLysCysAlaAspGlyLysV 397

    2156 TGTGTTTAGATCGGAAGTGCCTACAAATT 2184
         ::|||   :::|||:::|||:::::::::
     397 alCysLysAsnArgGlnCysValAspVal 406

seq_name: pir2:I52361

seq_documentation_block:
testicular metalloproteinase-like, disintegrin-like, cysteine-rich protein IVa - crab-ea
C;Species: Macaca fascicularis (crab-eating macaque)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000
C;Accession: I52361
[redacted]erry, A.C.F.; Jones, R.; Hall, L.
[redacted]hem. J. 312, 239-244, 1995
[redacted]le: Analysis of transcripts encoding novel members of the mammalian metalloprotease
[redacted]ve monkey tissues.
A;Reference number: I52361; MUID:96077150
A;Accession: I52361
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-732 <RES>
A;Cross-references: EMBL:X87205; NID:g1061158; PIDN:CAA60663.1; PID:g1061159
C;Genetics:
A;Gene: tMDC IVa
C;Superfamily: mouse meltrin alpha; disintegrin homology
F;403-485/Domain: disintegrin homology <DIS>


alignment_scores:
            Quality:  572.50              Length:    662
              Ratio:  1.608                Gaps:     35
 Percent Similarity:  53.776   Percent Identity: 27.341

alignment_block:
 US-09-634-252A-2 x I52361   ..

 Align seg 1/1  to: I52361  from: 1  to: 732

     586 TGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCT 635
         |||||||||   |||   ::::::   |||      |||   |||:::::
     109 CysTyrTyrLeuGlyTyrLeuGluGluValProLeuSerMetValThrVa 125

     636 GTCAACCTGC...AATGGACTTCATGGCATGTTTGAAGATGATACCTTCG 682
         ::::||||||   :::||||||:::|||:::::::::   |||   ::::
     125 lAspThrCysTyrGlyGlyLeuArgGlyIleMetLysLeuAspAspLeuA 142

     683 TGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACAGGT 732
         ::|||   |||:::||||||                ::::   |||
     142 laTyrGluIleLysProLeu..............GlnAspSerArgArg 153

     733 CGACCACATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAAT 782
               |||:::::::::::   :::|||         :::   :::
     154 PheGluHisValValSerGlnIleValAla.........GluProAsnAl 167

     783 GAAGAATCTCACTATGGAAAGAGGTGACCAGTGG.........CCCTTTC 823
             :::   |||::::::   ||||||::            |||::::
     167 aThrGlyProThrPheArgAspGlyAspAsnGluGluThrAspProLeuP 184

     824 TCTCTGAATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCA 873
         ::||||||                 :::   ::::::||||||
     184 heSerGlu.................AlaAsnAspSerMetAsnProArg 194

     874 CGTGGTATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAATGA 923
                                  ::::::   ::::::   :::::
     195 ........................IleSerThrLeuLeuTyrSerSe 202

     924 TCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAACAAC. 972
         :|||:::         |||         :::|||::::::   :::|||
     202 rHisArgGlyAsnIleLys.........GlyHisValGlnCysSerAsnT 216

     973 ....................................TTTGCAAAG 981
                                             :::
     216 hrTyrTyrArgIleTyrGlyAsnIleThrThrCysTyrIleAlaValVal 232

     982 TCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAG 1031
         ::::::   :::::::::|||||||||      ::::::::::   ||
     233 GlnMetPheSerValIleAspSerIleVal...GlnAsnIleAspLeuAr 248

    1032 GGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACA 1081
         |      :::      :::      ::::::   :::|||         :
     248 gTyrTyrIleTyrLeuLeuThrIleTyrAsnValArgAspProAlaProV 265

    1082 TCACC.........ACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAA 1122
         :::::         :::|||   :::::::::   :::
     265 alAsnGlnTyrGlyAlaAsnSerAlaMetPheThrTyrPheLysArgThr 281

    1123 TACCGGCAGCGCATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCG 1172
         :::   :::   ::::::   |||:::::::::::::   ||||||:::::
     282 PhePheAspThrPheArgValHisSerSerThrLeu...LeuIleLysGl 297

    1173 GGTGACATTT......CACTATAAGAGAAGCAGTCTGAGTTACTTTGGAG 1216
         ::::            :::|||:::   :::      |||:::
     297 uAlaProHisGluSerAsnTyrGluProGlnArgTyrSerPhe....... 311

    1217 GTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCA 1266
              |||::::::          :::   ::::::      |||
     312 .....CysThrHisLeuGlyLeuLeuHisIleGlyThrLeuGlyArgHis 326

    1267 ...ATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGG 1313
            :::   ||||||   :::   :::|||:::|||   ::::::::::||
     327 TyrLeuLeuValAlaValIleThrThrGlnThrLeuMetArgSerMetGl 343

    1314 AATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAAT 1363
         |::::::::::::   :::               |||   |||   :::
     343 yValGluTyrAspAspAsn............TyrCysThrCys...GlnA 355

    1364 CCTGGGGTGGCTGCATCATGGAGGAA...ACAGGGGTGTCCCATTCTCGA 1410
              :::   ||||||||||::::::      |||::::::   :::
     355 rgArgAlaPheCysIleMetGlnArgTyrProGlyIleThrAspAla... 370

    1411 AAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGG 1460
            ||||||:::||||||      :::   ::::::
     371 ...PheSerAsnCysSerTyrGlyHisThrGlnAsn.............. 381

    1461 AGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCC.... 1506
                    |||:::   |||             :::||||||
     382 ..........CysPheValAsnSerSerGlnCysValPheGluAlaLeuA 395

    1507 ............................ACGGAATGTGGAAATGGATAC 1527
                                      ::::::||||||||
     395 laProValHisAsnGluSerIleThrMetValArgCysGlyAsnLeuIle 411

    1528 GTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTAT.. 1575
         ||||||:::   ||||||||||||||||||         :::||||||
     412 ValGluGlyArgGluGluCysAspCysGlySerPheLysGlnCysTyrAl 428

    1576 .GGATTATGCTGTAAGAAA...TGTTCCCTCTCCAACGGGGCTCACTGCA 1621
          :::   ||||||::::::   |||   |||:::   |||:::   |||
     428 aSerArgCysCysArgSerAspCysArgLeuThrProGlySerIleCysH 445

    1622 GCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGG 1671
              |||   ||||||      |||:::|||   |||:::   :::|||
     445 isLeuGlyAspCysCys......ThrAsnCysSerPheSerAlaGlnGly 459

    1672 TATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTAC 1721
               ||||||         |||   |||||||:::   ||||||||||
     460 ThrLeuCysArgProIleGlnAsnIleCysAspLeuProGluTyrCysHi 476

    1722 TGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATG 1771
```

```
      |||   :::      |||||||||:::|||:::   |||||||||
  476 sGlyThrThrLeuThrCysProProAspLeuTyrLeuGlnAspGlyThrP 493

 1772 CATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGA 1821
        |||::::::   :::|||   ||||||:::|||:::|||      |||
  493 roCysThrGlu...GluGlyTyrCysTyrHisGlyAsnCysThrAspArg 508

 1822 GACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAA 1871
      :::      |||:::   |||:::|||:::::::|||      :::
  509 AsnValLeuCysLysAlaIlePheGlyValSerAlaLysAspAlaProGl 525

 1872 GTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCG 1921
        ||||||:::   :::|||   |||:::   :::   |||:::|||
  525 yAspCysTyrAsp...IleAsnLeuGluSerHisArgPheGlyHisCysT 541

  922 GGAAGGATGGAGAC......CGGTGGATTCAGTGCAGCAAACATGATGTG 1965
        ::::::   :::         :::      |||         |||
  541 hrArgGluGlnSerAlaLeuSerTyrGlnAlaCysValGlyIleAspLys 557

 1966 TTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGG 2015
      |||||||||   |||   |||||||||:::|||:::   ||||||:::
  558 PheCysGlyArgLeuGlnCysThrAsnValThrHisLeuProArgLeuGl 574

 2016 TCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAA...GGCC 2062
        :::                  :::||||||:::|||:::      |
  574 nGlu.................HisValSerPheHisHisSerIleArgA 585

 2063 GGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGAT..... 2100
      ||   ::::::|||   |||            ||||||
  585 rgGlyPheGlnCysPheGly...........LeuAspGluHisArgAla 597

 2101 ...GATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCC 2147
         |||   ||||||||||||   |||   ||||||||||||||||:::
  598 ThrAspThrThrAspValGlyArgValIleAspGlyThrProCysSerAr 514

 2148 GTCTATGATGTGTTTAGATCGGAAGTGC...CTACAAATTCAAGCCCTAA 2194
        :::|||   |||   :::   :::|||   :::   |||   :::||||
  614 gGlyMetAlaCysAsnAsnThrGlnCysAsnValAlaIleThrSerLeuA 531

 2195 ATATGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCAT 2244
      ||   :::      |||               :::   ||||||
  631 snTyrAspCysTyrPro.............GlnLysCysSerTyrLys 542

 2245 GGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGG 2294
      |||||||||:::|||:::   :::|||   |||   :::   |||
  643 GlyValCysAsnAsnArgArgAsnCysHisCysHisIleGlyTrpAspPr 559

 2295 GACAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAAGG 2344
            |||   :::|||          |||::::::          |
  659 oProArgCys...LeuArg.........ArgGlyIleGlyGlySerValA 672

 2345 ATGAAGGACCCAAGGGTCCTAGTGCCACCAATCTCATAATAGGCTCCATC 2394
      ||:::||||||      |||         :::      :::   :::
  672 spSerGlyPro......ProProArgArgThrArgSerValLysGlnSer 686

 2395 GCTGGTGCCATCCTGGTAGCAGCTATTGTCCTTGGG 2430
            |||:::|||         :::|||:::|||
  687 HisGlnAlaValLeuTyrLeuArgValValPheGly 698
```

seq_name: pir2:I65253

seq_documentation_block:
disintegrin-like testicular metalloproteinase (EC 3.4.24.-) IVb - crab-eating macaque (f
C;Species: Macaca fascicularis (crab-eating macaque)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C;Accession: I65253
R;Perry, A.C.F.; Jones, R.; Hall, L.
Biochem. J. 312, 239-244, 1995
A;Title: Analysis of transcripts encoding novel members of the mammalian metalloprotease
uctive monkey tissues.
A;Reference number: I52361; MUID:96077150
A;Accession: I65253
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-713 <RES>
A;Cross-references: EMBL:X87206; NID:g1061160; PIDN:CAA60664.1; PID:g1061161
C;Genetics:
A;Gene: tMDC IVb
C;Superfamily: mouse meltrin alpha; disintegrin homology
C;Keywords: hydrolase; metalloproteinase
F;384-466/Domain: disintegrin homology <DIS>

alignment_scores:
Quality: 566.50 Length: 667
Ratio: 1.587 Gaps: 38
Percent Similarity: 53.523 Percent Identity: 27.586

alignment_block:
US-09-634-252A-2 x I65253 ..

Align seg 1/1 to: I65253 from: 1 to: 713

```
  586 TGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCT 635
      |||||||||   |||   ::::::   |||       |||   |||:::::
   90 CysTyrTyrLeuGlyTyrLeuGluGluValProLeuSerMetValThrVa 106

  636 GTCAACCTGC...AATGGACTTCATGGCATGTTTGAAGATGATACCTTCG 682
      ::::||||||   :::|||||:::|||:::::::::   |||   ::::
  106 lAspThrCysTyrGlyGlyLeuArgGlyIleMetLysLeuAspAspLeuA 123

  683 TGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACAGGT 732
      ::|||   |||:::||||||                ::   |||
  123 laTyrGluIleLysProLeu..............GlnAspSerArgArg 134

  733 CGACCACATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAAT 782
            |||::::::::::::   :::|||          :::   :::
  135 PheGluHisValValSerGlnIleValAla.........GluProAsnAl 148

  783 GAAGAATCTCACTATGGAAAGAGGTGACCAGTGG.........CCCTTTC 823
          :::   |||:::::::   ||||||:::              |||::::
  148 aThrGlyProThrPheArgAspGlyAspAsnGluGluThrAspProLeuP 165

  824 TCTCTGAATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCA 873
      ::|||||                        :::   ::::::||||||
  165 heSerGlu..................AlaAsnAspSerMetAsnProArg 175

  874 CGTGGTATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAATGA 923
                                                      ::
  176 Leu.....................................................Se 177

  924 TCACAAAACGTATAAGAAGCATCGCTCTTCT.........CATGCACATA 964
      :   :::   |||::::::||||||::::::          |||::::::
  177 rProSerLeuTyrSerSerHisArgGlyAsnIleLysGlyHisValGlnC 194

  965 CCAACAAC.................................TTTGCAAAG 981
        :::|||                                    :::|||
  194 ysSerAsnSerTyrTyrArgIleTyrAspAsnIleThrThrCysSerLys 210

  982 TCCGTGGTC.........AACCTTGTGGATTCTATTTACAAGGAGCAGCT 1022
      :::||||||         :::::::::||||||||      ::::::::
  211 GluValValGlnMetPheSerValIleAspSerIleVal...GlnAsnIl 226

 1023 CAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATC 1072
      ::::   |||      :::      :::      ::::::::::::|||
  226 eAspLeuArgTyrTyrIleTyrLeuLeuThrIleTyrAsnSerArgAspP 243

 1073 AGATTGACATCACC............ACCAACCCTGTGCAGATGCTCCAT 1110
              ::::::            :::|||:::      |||:::
  243 roAlaProValAsnAspTyrArgTyrArgSerProIle......LeuAsn 257

 1111 GAGTTCTCAAAA...TACCGGCAGCGCATTAAGCAGCATGCT........ 1149
         |||   :::   :::   :::   ::::::   |||:::
  258 TyrPheArgArgThrPhePheAspThrPheArgValHisSerSerThrLe 274

 1150 .........GATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATA 1189
```

```
                ||||||   |||                    :::|||:
 274 uLeuIleLysAspAlaProHis.............GluSerAsnTyrG 286

1190 AGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGA 1239
     ::           |||:::           |||||||:::
 286 luProArgArgTyrSerPhe...........CysSerHisLeuSerLeu 298

1240 GTTGGTGTG...........AATGAGTATGGTCTTCCAATGGCAGTGGC 1277
     :::   :::           ::::::|||        :::   |||||
 299 IleHisIleGlyThrLeuAspArgHisTyr.........LeuLeuValAl 312

1278 ACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAAC 1327
     |   :::   ::::::  |||   ::::::||||||:::   ::::::
 312 aValIleThrThrHisIleLeuMetArgSerLeuGlyValAlaTyrAsp. 328

1328 CTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGC 1377
                    |||  |||   :::         :::|||
 329 ...........TyrLysTyrCysThrCys...GlnArgArgThrSerCys 340

1378 ATCATGGAGGAA...ACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTG 1424
     ||||||:::::      |||:::::   :::      ||||||:::||
 341 IleMetGlnArgTyrProGlyMetThrAspAla......PheSerAsnCy 355

1425 CAGCATTTTGGAGTATAGAGAC...TTTTTACAGAGAGGAGGTGGAGCCT 1471
     ||||    :::   ::::::  |||::::::         |||   |
 355 sSerTyrGlyHisIleGlnAsnCysPheValSerThr......GlyArgC 370

1472 GCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAA..... 1512
     ||:::|||:::      |||  :::::::::  |||
 370 ysIlePheGluAlaLeuThrProValTyrAsnGluThrMetThrThrVal 386

1513 ...TGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTT 1559
        |||||||||      ||||||:::  ||||||||||||||||||
 387 LeuCysGlyAsnLeuIleValGluGlyLysGluGluCysAspCysGlySe 403

1560 TCATGTGGAATGCTAT...GGATTATGCTGTAAGAAA...TGTTCCCTCT 1603
          :::||||||  :::  ||||||:::::  |||  ::::
 403 rPheLysGlnCysTyrAlaAsnHisCysCysGlnSerAspCysHisPheT 420

1604 CCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGT 1653
     ::   |||:::  |||     |||  ||||||     |||:::|||
 420 hrProGlySerIleCysHisLeuGlyAspCysCys......ThrAsnCys 434

1654 CTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGA 1703
        |||:::  :::|||      ||||||         |||  |||||
 435 SerPheSerAlaGlnGlyThrLeuCysArgProIleGlnAsnIleCysAs 451

1704 TATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTC 1753
     |:::  |||||||||  |||  :::      |||||||||:::::::
 451 pLeuProGluTyrCysTyrGlyThrThrLeuThrCysProProAspPheT 468

1754 ATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAAT 1803
     ::  |||||||||      |||:::::  :::|||  ||||||:::
 468 yrLeuGlnAspGlyThrProCysThrGlu...GluGlyTyrCysTyrHis 483

1804 GGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAA 1853
     |||:::|||      |||:::      |||:::   |||:::|||:::::
 484 GlyAsnCysThrAspArgAsnValLeuCysLysAlaIlePheGlyValSe 500

1854 GGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCA 1903
     :|||      :::   :::  ||||||:::  :::|||  |||:::
 500 rAlaGluAspAlaProGluAspCysTyrAsp...IleAsnLeuGluAsnH 516

1904 CTGAGAAGGGAAACTGC......GGGAAGGATGGAGACCGGTGGATTCAG 1947
       :::  |||:::|||      ::::::  :::      :::
 516 isArgPheGlyHisCysThrArgAlaArgThrAlaIleAlaTyrGluAla 532

1948 TGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTAC 1997
     |||:::      |||  |||||||||  |||  |||||||||:::||
 533 CysAlaLeuIleAspLysPheCysGlyArgLeuGlnCysThrAsnValTh 549

1998 TCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCT 2047
     |:::  ||||||:::  :::                  :::||||
```

```
 549 rHisLeuProArgLeuGlnGlu.................HisValSerP 560

2048 TCTACCATCAAGGCCGG...GTGATTGACTGCAGTGGTGCCCATGTAGTT 2094
     ||:::|||:::  |||     :::::::|||  |||
 560 heHisHisSerValArgGlyGlyPheGlnCysPheGly............ 572

2095 TTAGAT..............GATGATACGGATGTGGGCTATGTAGAAGA 2129
     ||||||              |||  ||||||||||||:::|||  ||
 573 LeuAspGluHisArgAlaThrAspThrThrAspValGlyHisValIleAs 589

2130 TGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTAC 2179
     ||||||||||||||:::  ::::::::::|||  :::  :::|||
 589 pGlyThrProCysAlaAspGlyIlePheCysAsnAsnSerGlnCys.... 604

2180 AAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAAA 2229
                   ::::::::::|||        |||        :::
 605 ........AsnAlaThrIleThrSerLeuGlyTyrAspCysHisProGlu 618

2230 GTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGA 2279
        ||||||  :::||||||||||:::|||:::  :::|||  |||
 619 LysCysSerHisArgGlyValCysAsnAsnArgArgAsnCysHisCysHi 635

2280 TTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGTTAGGAACC 2329
      :::  |||          |||      |||          ::::
 635 sIleGlyTrpAspProProLeuCysLeuArgArgGlyAlaGlyGlySerV 652

2330 TTCACCCCCCCAAGGATGAAGGACCCAAGGGTCCTAGTGCCACCAATCTC 2379
     ::          |||:::||||||      |||        :::
 652 al...........AspSerGlyPro......ProProArgArgThrArg 662

2380 ATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCTATTGTCCTTGG 2429
        :::  :::         ::::::|||         :::|||:::||
 663 SerValLysGlnSerGlnGlnSerValLeuTyrLeuArgValValPheGl 679

2430 G 2430
     |
 679 y 679
```

```
OM of: US-09-634-252A-2 to: SPTREMBL_16:*   out_format : pfs

Date: Aug 15, 2001 10:01 PM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=xlp
 -Q=/cgn2_1/USPTO_spool/US09634252/runat_14082001_142629_9841/app_query.fasta_1.2598
 -DB=SPTREMBL_16 -QFMT=fastan -SUFFIX=n2p.rspt -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -[redacted]ANS=human40.cdi -LIST=90 -DOCALIGN=200 -THR_SCORE=pct
 -[redacted]_MAX=100 -THR_MIN=0 -ALIGN=30 -MODE=LOCAL -OUTFMT=pfs
 -[redacted]M=ext -MINLEN=0 -MAXLEN=2000000000
 -[redacted]SER=US09634252_@CGN1_1_168 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY
 -WAIT -THREADS=1

Search information block:
Query: US-09-634-252A-2
Query length: 2499
Database: SPTREMBL_16:*
Database sequences: 425026
Database length: 132305027
Search time (sec): 153.920000

score_list:
Sequence              Strd Orig    ZScore    EScore Len  ! Documentation ..
sp_human:O75077       +  4553.00 6247.23   0.0       832  ! O75077 homo sapiens (human). mo
sp_rodent:Q9R1V7      +  4259.50 5843.46   0.0       829  ! Q9r1v7 mus musculus (mouse). ac
sp_rodent:Q9R1V4      +  2027.00 2772.54   4.3e-147  773  ! Q9r1v4 mus musculus (mouse). ac
sp_human:Q9UIA1       +  1985.50 2714.45   6.6e-144  870  ! Q9uia1 homo sapiens (human). mo
sp_human:Q9UKK2       +  1985.00 2713.79   7.2e-144  867  ! Q9ukk2 homo sapiens (human). me
sp_human:O75076       +  1983.50 2712.16   9.4e-144  823  ! O75076 homo sapiens (human). mo
sp_human:O75075       +  1983.50 2711.80   9.4e-144  859  ! O75075 homo sapiens (human). mo
sp_human:Q9P0K2       +  1983.50 2711.42   9.4e-144  899  ! Q9p0k2 homo sapiens (human). me
sp_human:Q9P0K1       +  1983.50 2711.35   9.4e-144  906  ! Q9p0k1 homo sapiens (human). me
sp_human:O75078       +  1983.00 2712.05   1.0e-143  769  ! O75078 homo sapiens (human). mo
sp_rodent:Q9R1V5      +  1971.00 2694.99   8.5e-143  821  ! Q9r1v5 mus musculus (mouse). ac
sp_rodent:Q9R1V6      +  1971.00 2694.62   8.5e-143  857  ! Q9r1v6 mus musculus (mouse). ac
sp_vertebrate:O42596  +  1944.00 2656.74   1.0e-140  935  ! O42596 xenopus laevis (african
sp_human:Q14809       +  1868.50 2555.76   6.0e-135  664  ! Q14809 homo sapiens (human). me
sp_vertebrate:Q9PSZ3  +  1453.00 1987.35   4.0e-103  452  ! Q9psz3 xenopus laevis (african
sp_human:Q14810       +  1298.00 1772.95   3.1e-91   518  ! Q14810 homo sapiens (human). me
sp_human:Q14808       +  1297.00 1771.48   3.7e-91   524  ! Q14808 homo sapiens (human). mo
sp_human:Q9H013       +  1011.50 1373.95   2.9e-69   918  ! Q9h013 homo sapiens (human). me
sp_rodent:O35674      +   975.00 1323.71   1.8e-66   920  ! O35674 mus musculus (mouse). me
sp_vertebrate:O12960  +   968.50 1314.82   5.7e-66   914  ! O12960 xenopus laevis (african
sp_rodent:Q61824      +   966.00 1311.48   8.9e-66   903  ! Q61824 mus musculus (mouse). me
sp_human:Q13443       +   920.50 1249.71   2.7e-62   819  ! Q13443 homo sapiens (human). ce
sp_human:Q9UHP2       +   916.00 1243.14   6.0e-62   857  ! Q9uhp2 homo sapiens (human). di
sp_rodent:Q61072      +   892.00 1210.23   4.1e-60   845  ! Q61072 mus musculus (mouse). ce
sp_invertebrate:Q9VXL1 +  883.00 1199.63   2.0e-59   685  ! Q9vxl1 drosophila melanogaste
sp_invertebrate:Q9GZ15 +  883.00 1199.52   2.0e-59   694  ! Q9gz15 drosophila melanogaste
sp_vertebrate:O42595  +   866.50 1174.88   3.7e-58   873  ! O42595 xenopus laevis (african
sp_invertebrate:Q9VFH0 +  850.00 1147.22   7.2e-57  1569  ! Q9vfh0 drosophila melanogast
sp_mammal:Q9XSL6      +   842.00 1142.16   2.8e-56   776  ! Q9xsl6 macaca fascicularis (cra
sp_human:Q13493       +   826.00 1119.75   4.7e-55   814  ! Q13493 homo sapiens (human). mo
sp_human:Q13444       +   825.00 1118.37   5.6e-55   814  ! Q13444 homo sapiens (human). me
sp_human:Q9H2U9       +   815.50 1105.95   3.0e-54   754  ! Q9h2u9 homo sapiens (human). a
sp_human:Q9Y3S0       +   812.50 1101.59   5.1e-54   775  ! Q9y3s0 homo sapiens (human). em
sp_human:Q9UKQ2       +   810.50 1098.84   7.2e-54   775  ! Q9ukq2 homo sapiens (human). me
sp_rodent:Q9JI76      +   799.00 1083.53   5.5e-53   729  ! Q9ji76 mus musculus (mouse). me
sp_mammal:Q28475      +   797.50 1080.94   7.2e-53   776  ! Q28475 macaca fascicularis (cra
sp_mammal:O19061      +   790.50 1071.14   2.5e-52   792  ! O19061 saguinus oedipus (cotton
sp_rodent:Q9QYV0      +   788.00 1067.44   3.9e-52   816  ! Q9qyv0 rattus norvegicus (rat).
sp_rodent:O88839      +   783.00 1060.58   9.3e-52   815  ! O88839 mus musculus (mouse). me
sp_mammal:O19060      +   783.00 1060.51   9.3e-52   821  ! O19060 saguinus oedipus (cotton
sp_mammal:O19056      +   780.50 1056.90   1.5e-51   838  ! O19056 papio anubis (olive b
sp_mammal:O46652      +   780.00 1056.35   1.6e-51   825  ! O46652 papio anubis (olive b
sp_mammal:Q28477      +   779.00 1054.97   1.9e-51   825  ! Q28477 macaca fascicularis (
sp_vertebrate:Q90499  +   778.50 1056.74   2.0e-51   617  ! Q90499 echis pyramidum (carp
sp_rodent:Q63180      +   775.00 1049.84   3.8e-51   789  ! Q63180 rattus norvegicus (ra
sp_mammal:Q28476      +   775.00 1048.68   3.9e-51   905  ! Q28476 macaca fascicularis (
sp_rodent:Q9QYL2      +   772.00 1045.45   6.5e-51   814  ! Q9qyl2 mus musculus (mouse).
sp_mammal:O19057      +   769.50 1041.79   1.0e-50   836  ! O19057 pongo pygmaeus (orang
sp_human:Q9UKJ9       +   765.00 1036.79   2.2e-50   726  ! Q9ukj9 homo sapiens (human).
sp_human:Q9UKJ8       +   760.00 1029.96   5.3e-50   722  ! Q9ukj8 homo sapiens (human).
sp_rodent:Q9JLN6      +   759.50 1028.68   5.9e-50   774  ! Q9jln6 mus musculus (mouse).
sp_vertebrate:Q9PT48  +   756.50 1026.65   9.8e-50   604  ! Q9pt48 atractaspis engaddens
sp_mammal:Q28478      +   754.50 1022.24   1.4e-49   735  ! Q28478 macaca fascicularis (
sp_rodent:O35227      +   754.00 1020.96   1.6e-49   788  ! O35227 mus musculus (mouse).
sp_rodent:P70505      +   751.00 1016.82   2.6e-49   789  ! P70505 rattus norvegicus (ra
sp_human:O43506       +   749.00 1014.77   3.7e-49   726  ! O43506 homo sapiens (human).
sp_vertebrate:Q9PVK7  +   748.00 1015.01   4.4e-49   600  ! Q9pvk7 naja naja (indian cob
sp_vertebrate:Q9YI20  +   748.00 1014.87   4.4e-49   610  ! Q9yi20 agkistrodon halys pal
sp_invertebrate:P90974 +  746.50 1008.28   6.0e-49  1042  ! P90974 caenorhabditis ele
sp_human:Q9UQQ8       +   743.00 1006.42   1.1e-48   735  ! Q9uqq8 homo sapiens (human).
sp_mammal:Q28659      +   737.50  996.96   2.9e-48   919  ! Q28659 oryctolagus cuniculus
sp_vertebrate:Q90495  +   735.00  996.90   4.3e-48   616  ! Q90495 echis carinatus (saw-
sp_vertebrate:O93523  +   729.00  988.73   1.3e-47   610  ! O93523 bothrops jararaca (ja
sp_mammal:O77779      +   727.50  984.25   1.7e-47   812  ! O77779 bos taurus (bovine).
sp_vertebrate:O42138  +   726.50  985.15   1.9e-47   620  ! O42138 agkistrodon contortri
sp_vertebrate:Q9W6M5  +   725.00  983.24   2.5e-47   609  ! Q9w6m5 agkistrodon acutus (d
sp_vertebrate:Q9DGB9  +   725.00  983.23   2.5e-47   610  ! Q9dgb9 crotalus atrox (weste
sp_rodent:Q9R159      +   725.00  981.37   2.6e-47   760  ! Q9r159 mus musculus (mouse).
sp_vertebrate:Q90282  +   723.50  981.18   3.3e-47   609  ! Q90282 crotalus atrox (weste
sp_human:Q9UKF2       +   720.00  974.16   6.3e-47   790  ! Q9ukf2 homo sapiens (human).
sp_human:Q9UKF1       +   719.00  972.88   7.5e-47   781  ! Q9ukf1 homo sapiens (human).
sp_rodent:Q9R158      +   718.50  973.16   8.1e-47   697  ! Q9r158 mus musculus (mouse).
sp_rodent:Q60410      +   717.50  970.57   9.7e-47   804  ! Q60410 cavia porcellus (guin
sp_human:Q9UKF4       +   717.00  970.30   1.1e-46   766  ! Q9ukf4 homo sapiens (human).
sp_human:Q9UKF3       +   717.00  970.07   1.1e-46   787  ! Q9ukf3 homo sapiens (human).
sp_human:Q9UKF5       +   717.00  969.72   1.1e-46   820  ! Q9ukf5 homo sapiens (human).
sp_rodent:Q60411      +   716.00  969.27   1.3e-46   735  ! Q60411 cavia porcellus (guin
sp_human:Q9UHP1       +   715.00  966.97   1.5e-46   820  ! Q9uhp1 homo sapiens (human).
sp_mammal:Q28660      +   713.00  964.96   2.1e-46   751  ! Q28660 oryctolagus cuniculus
sp_mammal:O77780      +   705.00  954.02   8.8e-46   745  ! O77780 bos taurus (bovine).
sp_rodent:Q9R160      +   704.50  953.15   9.6e-46   761  ! Q9r160 mus musculus (mouse).
sp_rodent:Q63202      +   702.50  950.65   1.4e-45   739  ! Q63202 rattus norvegicus (ra
sp_rodent:Q9R157      +   697.50  944.00   3.3e-45   719  ! Q9r157 mus musculus (mouse).
sp_human:Q9Y3Q7       +   696.50  942.39   3.9e-45   739  ! Q9y3q7 homo sapiens (human).
sp_mammal:Q95194      +   693.50  938.19   6.7e-45   746  ! Q95194 macaca fascicularis (
sp_rodent:Q9QWJ0      +   691.00  934.87   1.0e-44   735  ! Q9qwj0 mus musculus (mouse).
sp_human:O43507       +   689.50  935.38   1.3e-44   542  ! O43507 homo sapiens (human).
sp_vertebrate:O42593  +   689.00  932.46   1.5e-44   706  ! O42593 xenopus laevis (afric
sp_rodent:Q60472      +   687.00  928.90   2.1e-44   777  ! Q60472 cavia porcellus (guin
sp_rodent:Q60718      +   685.00  926.62   3.0e-44   735  ! Q60718 mus musculus (mouse).

seq_name: sp_human:O75077

seq_documentation_block:
ID   O75077      PRELIMINARY;      PRT;   832 AA.
AC   O75077;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MDC3 (ADAM22 PROTEIN).
GN   ADAM22.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=98359734; PubMed=9693107;
RA   Sagane K., Ohya Y., Hasegawa Y., Tanaka I.;
RT   "Metalloproteinase-like, disintegrin-like, cysteine-rich proteins MDC2
RT   and MDC3: novel human cellular disintegrins highly expressed in the
RT   brain.";
```

```
RL   Biochem. J. 334:93-98(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RA   Cal S., Lopez-Otin C.;
RT   "Identification and characterization of ADAM 22, a novel
RT   metalloprotease/disintegrin protein.";
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB009672; BAA32351.1; -.
DR   EMBL; AJ005580; CAC20565.1; -.
DR   HSSP; P18619; 1FVL.
DR   MEROPS; M12.979; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
     ProDom; PD000664; -; 1.
     PROSITE; PS50215; ADAM_MEPRO; 1.
     PROSITE; PS50214; DISINTEGRIN_2; 1.
     PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   SMART; SM00050; DISIN; 1.
SQ   SEQUENCE   832 AA;  91925 MW;  7841A9670E1C24EF CRC64;
```

```
alignment_scores:
            Quality: 4553.00              Length:    832
              Ratio:  5.472                Gaps:      0
 Percent Similarity: 100.000   Percent Identity: 100.000

alignment_block:
 US-09-634-252A-2 x O75077   ..

 Align seg 1/1  to: O75077  from: 1  to: 832

       1 ATGAAGCCGCCCGGCAGCAGCTCGCGGCAGCCGCCCCTGGCGGGCTGCAG 50
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 MetLysProProGlySerSerSerArgGlnProProLeuAlaGlyCysSe 17

      51 CCTTGCCGGCGCTTCCTGCGGCCCCCAACGCGGCCCCGCCGGCTCGGTGC 100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 rLeuAlaGlyAlaSerCysGlyProGlnArgGlyProAlaGlySerValP 34

     101 CTGCCAGCGCCCCGGCCCGCACGCCGCCCTGCCGCCTGCTTCTCGTCCTT 150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 roAlaSerAlaProAlaArgThrProProCysArgLeuLeuLeuValLeu 50

     151 CTCCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCGCGCCTGGGGGGC 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 LeuLeuLeuProProLeuAlaAlaSerSerArgProArgAlaTrpGlyAl 67

     201 TGCTGCGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAAAAAAATTTGG 250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      67 aAlaAlaProSerAlaProHisTrpAsnGluThrAlaGluLysAsnLeuG 84

     251 GAGTCCTGGCAGATGAAGACAATACATTGCAACAGAATAGCAGCAGTAAT 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      84 lyValLeuAlaAspGluAspAsnThrLeuGlnGlnAsnSerSerSerAsn 100

     301 ATCAGTTACAGCAATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACT 350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     101 IleSerTyrSerAsnAlaMetGlnLysGluIleThrLeuProSerArgLe 117

     351 CATATATTACATCAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACA 400
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     117 uIleTyrTyrIleAsnGlnAspSerGluSerProTyrHisValLeuAspT 134

     401 CAAAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAG 450
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     134 hrLysAlaArgHisGlnGlnLysHisAsnLysAlaValHisLeuAlaGln 150

     451 GCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCAT 500
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     151 AlaSerPheGlnIleGluAlaPheGlySerLysPheIleLeuAspLeuIl 167

     501 ACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAA 550
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     167 eLeuAsnAsnGlyLeuLeuSerSerAspTyrValGluIleHisTyrGluA 184

     551 ATGGGAAACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGA 600
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     184 snGlyLysProGlnTyrSerLysGlyGlyGluHisCysTyrTyrHisGly 200

     601 AGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGG 650
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     201 SerIleArgGlyValLysAspSerLysValAlaLeuSerThrCysAsnGl 217

     651 ACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCAC 700
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     217 yLeuHisGlyMetPheGluAspAspThrPheValTyrMetIleGluProL 234

     701 TAGAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCACATATAATCCAG 750
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     234 euGluLeuValHisAspGluLysSerThrGlyArgProHisIleIleGln 250

     751 AAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGA 800
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     251 LysThrLeuAlaGlyGlnTyrSerLysGlnMetLysAsnLeuThrMetGl 267

     801 AAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAA 850
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     267 uArgGlyAspGlnTrpProPheLeuSerGluLeuGlnTrpLeuLysArgA 284

     851 GGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATAT 900
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     284 rgLysArgAlaValAsnProSerArgGlyIlePheGluGluMetLysTyr 300

     901 TTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTC 950
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     301 LeuGluLeuMetIleValAsnAspHisLysThrTyrLysLysHisArgSe 317

     951 TTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGG 1000
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     317 rSerHisAlaHisThrAsnAsnPheAlaLysSerValValAsnLeuValA 334

    1001 ATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTA 1050
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     334 spSerIleTyrLysGluGlnLeuAsnThrArgValValLeuValAlaVal 350

    1051 GAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCA 1100
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     351 GluThrTrpThrGluLysAspGlnIleAspIleThrThrAsnProValGl 367

    1101 GATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATTAAGCAGCATGCTG 1150
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     367 nMetLeuHisGluPheSerLysTyrArgGlnArgIleLysGlnHisAlaA 384

    1151 ATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGT 1200
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     384 spAlaValHisLeuIleSerArgValThrPheHisTyrLysArgSerSer 400

    1201 CTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAA 1250
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     401 LeuSerTyrPheGlyGlyValCysSerArgThrArgGlyValGlyValAs 417

    1251 TGAGTATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTGG 1300
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     417 nGluTyrGlyLeuProMetAlaValAlaGlnValLeuSerGlnSerLeuA 434

    1301 CTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATGT 1350
         ||||||||||||||||||||||||||||||||||||||||||||||||||
     434 laGlnAsnLeuGlyIleGlnTrpGluProSerSerArgLysProLysCys 450

    1351 GACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTC 1400
```

```
451 AspCysThrGluSerTrpGlyGlyCysIleMetGluGluThrGlyValSe 467

1401 CCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTT 1450
467 rHisSerArgLysPheSerLysCysSerIleLeuGluTyrArgAspPheL 484

1451 TACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTT 1500
484 euGlnArgGlyGlyGlyAlaCysLeuPheAsnArgProThrLysLeuPhe 500

1501 GAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGA 1550
501 GluProThrGluCysGlyAsnGlyTyrValGluAlaGlyGluGluCysAs 517

551 TTGTGGTTTTCATGTGGAATGCTATGGATTATGCTGTAAGAAATGTTCCC 1600
517 pCysGlyPheHisValGluCysTyrGlyLeuCysCysLysLysCysSerL 534

1601 TCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCA 1650
534 euSerAsnGlyAlaHisCysSerAspGlyProCysCysAsnAsnThrSer 550

1651 TGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTG 1700
551 CysLeuPheGlnProArgGlyTyrGluCysArgAspAlaValAsnGluCy 567

1701 TGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATC 1750
567 sAspIleThrGluTyrCysThrGlyAspSerGlyGlnCysProProAsnL 584

1751 TTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTAC 1800
584 euHisLysGlnAspGlyTyrAlaCysAsnGlnAsnGlnGlyArgCysTyr 500

1801 AATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAAC 1850
601 AsnGlyGluCysLysThrArgAspAsnGlnCysGlnTyrIleTrpGlyTh 517

1851 AAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAG 1900
617 rLysAlaAlaGlySerAspLysPheCysTyrGluLysLeuAsnThrGluG 534

1901 GCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGC 1950
634 lyThrGluLysGlyAsnCysGlyLysAspGlyAspArgTrpIleGlnCys 550

951 AGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCG 2000
651 SerLysHisAspValPheCysGlyPheLeuLeuCysThrAsnLeuThrAr 667

2001 AGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCT 2050
667 gAlaProArgIleGlyGlnLeuGlnGlyGluIleIleProThrSerPheT 684

2051 ACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGAT 2100
684 yrHisGlnGlyArgValIleAspCysSerGlyAlaHisValValLeuAsp 700

2101 GATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTC 2150
701 AspAspThrAspValGlyTyrValGluAspGlyThrProCysGlyProSe 717

2151 TATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGA 2200
717 rMetMetCysLeuAspArgLysCysLeuGlnIleGlnAlaLeuAsnMetS 734

2201 GCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTG 2250
734 erSerCysProLeuAspSerLysGlyLysValCysSerGlyHisGlyVal 750

2251 TGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGA 2300
```

```
751 CysSerAsnGluAlaThrCysIleCysAspPheThrTrpAlaGlyThrAs 767

2301 TTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAAGGATGAAG 2350
767 pCysSerIleArgAspProValArgAsnLeuHisProProLysAspGluG 784

2351 GACCCAAGGGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGGT 2400
784 lyProLysGlyProSerAlaThrAsnLeuIleIleGlySerIleAlaGly 800

2401 GCCATCCTGGTAGCAGCTATTGTCCTTGGGGGCACAGGCTGGGGATTTAA 2450
801 AlaIleLeuValAlaAlaIleValLeuGlyGlyThrGlyTrpGlyPheLy 817

2451 AAATGTCAAGAAGAGAAGGTTCGATCCTACTCAGCAAGGCCCCATC 2496
817 sAsnValLysLysArgArgPheAspProThrGlnGlnGlyProIle 832
```

```
seq_name: sp_rodent:Q9R1V7

seq_documentation_block:
ID   Q9R1V7      PRELIMINARY;      PRT;   829 AA.
AC   Q9R1V7;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   ADAM23.
GN   ADAM23.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=99365303; PubMed=10433968;
RA   Sagane K., Yamazaki K., Mizui Y., Tanaka I.;
RT   "Cloning and chromosomal mapping of mouse ADAM11,ADAM22 and ADAM23.";
RL   Gene 236:79-86(1999).
DR   EMBL; AB009673; BAA83381.1; -.
DR   HSSP; P18619; 1FVL.
DR   MGD; MGI:1345162; Adam23.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   SMART; SM00050; DISIN; 1.
SQ   SEQUENCE   829 AA;  91547 MW;  F6EBCD69DD50F53A CRC64;

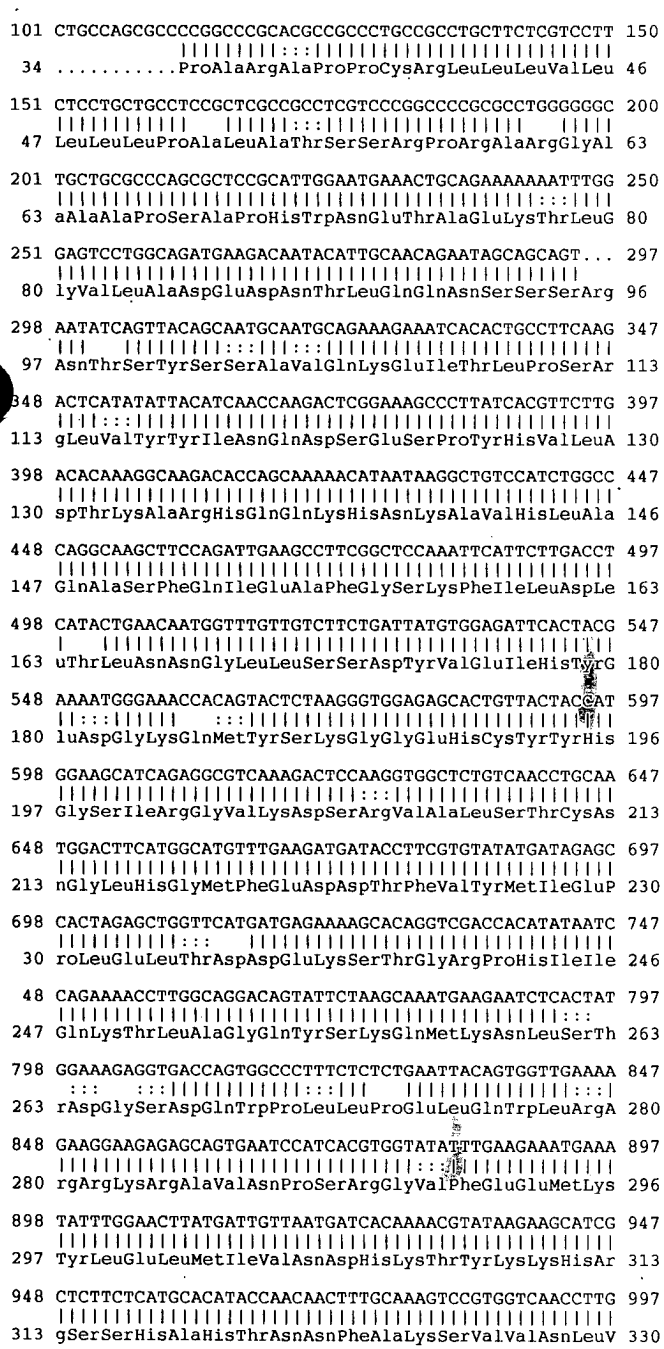

alignment_scores:
            Quality: 4259.50              Length:    833
              Ratio: 5.278                  Gaps:      2
 Percent Similarity: 96.879  Percent Identity: 93.758

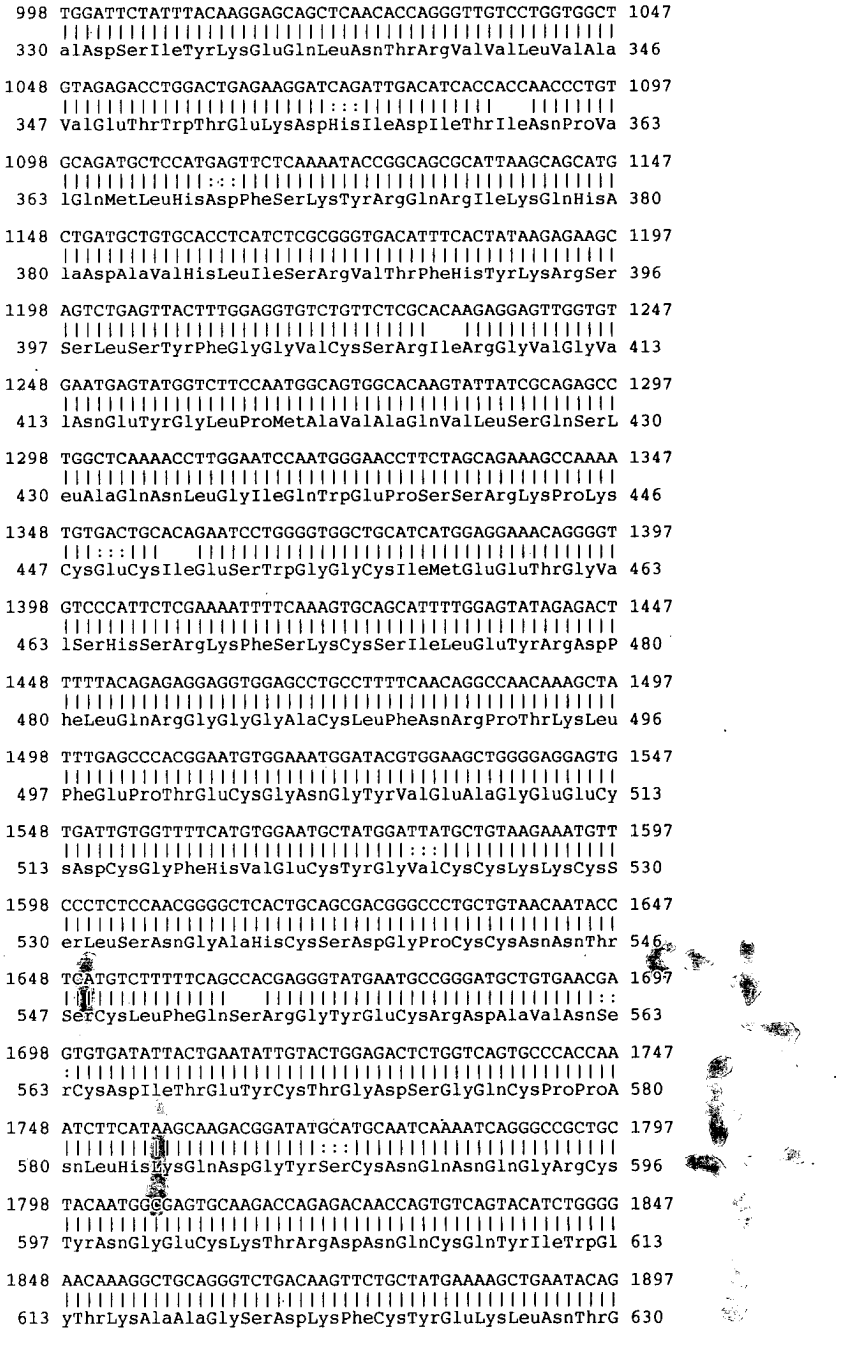
alignment_block:
 US-09-634-252A-2 x Q9R1V7   ..

 Align seg 1/1  to: Q9R1V7  from: 1  to: 829

       1 ATGAAGCCGCCCGGCAGCAGCTCGCGGCAGCCGCCCCTGGCGGGCTGCAG 50
         |||||||||||||||||   ||||||:::|||   |||:::|||||||||
       1 MetLysProProGlySerIleSerArgArgProThrLeuThrGlyCysSe 17

      51 CCTTGCCGGCGCTTCCTGCGGCCCCCAACGCGGCCCCGCCGGCTCGGTGC 100
         ||||   |||||||||||||||||   |||   |||||||||   |||
      17 rLeuProGlyAlaSerCysGlyProGlyArgCysProAlaGlyProVal. 33
```





```
     ||||||||||||||||||||||||||||||||||||||||||||||
 630 luGlyThrGluLysGlyAsnCysGlyLysAspGlyAspArgTrpIlePro 646

1948 TGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTAC 1997
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 647 CysSerLysHisAspValPheCysGlyPheLeuLeuCysThrAsnLeuTh 563

1998 TCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCT 2047
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 663 rArgAlaProArgIleGlyGlnLeuGlnGlyGluIleIleProThrSerP 680

2048 TCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTA 2097
     ||||||||||||||||||||||||||||||||||||||||||||||||||||
 680 heTyrHisGlnGlyArgValIleAspCysSerGlyAlaHisValValLeu 696

 098 GATGATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCC 2147
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 697 AspAspAspThrAspValGlyTyrValGluAspGlyThrProCysGlyPr 713

2148 GTCTATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATA 2197
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 713 oSerMetMetCysLeuAspArgLysCysLeuGlnIleGlnAlaLeuAsnM 730

2198 TGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGG 2247
     ||||||||||||||||||||:::||||||||||||||||||||||||||
 730 etSerSerCysProLeuAspSerArgGlyLysValCysSerGlyHisGly 746

2248 GTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGAC 2297
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 747 ValCysSerAsnGluAlaThrCysIleCysAspPheThrTrpAlaGlyTh 763

2298 AGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAAGGATG 2347
     ||||||||||||||||||||||||||||||   :::|||||||||||||
 763 rAspCysSerIleArgAspProValArgAsnProAsnProProLysAspG 780

2348 AAGGACCCAAGGGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCT 2397
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 780 luGlyProLysGlyProSerAlaThrAsnLeuIleIleGlySerIleAla 796

2398 GGTGCCATCCTGGTAGCAGCTATTGTCCTTGGGGGCACAGGCTGGGGATT 2447
     ||||||||||||||||||||||||||||||||||||||||||||||||||
 797 GlyAlaIleLeuValAlaAlaIleValLeuGlyGlyThrGlyTrpGlyPh 813

2448 TAAAAATGTCAAGAAGAGAAGGTTCGATCCTACTCAGCAAGGCCCCATC 2496
     ||||||||||||||||||||||||||||||||||||||||||||||||
 813 eLysAsnValLysLysArgArgPheAspProThrGlnGlnGlyProIle 829

Name: sp_rodent:Q9R1V4

Documentation_block:
ID   Q9R1V4      PRELIMINARY;      PRT;   773 AA.
AC   Q9R1V4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   ADAM11.
GN   ADAM11.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=99365303; PubMed=10433968;
RA   Sagane K., Yamazaki K., Mizui Y., Tanaka I.;
RT   "Cloning and chromosomal mapping of mouse ADAM11,ADAM22 and ADAM23.";
RL   Gene 236:79-86(1999).
DR   EMBL; AB009676; BAA83384.1; -.
DR   HSSP; P18619; 1FVL.
DR   MGD; MGI:1098667; Adam11.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PRINTS; PR00289; DISINTEGRIN.
DR   ProDom; PD000664; -; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   SMART; SM00050; DISIN; 1.
SQ   SEQUENCE   773 AA;  84134 MW;  9A8A125FB3F743DD CRC64;


alignment_scores:
           Quality: 2027.00              Length:    806
             Ratio:  3.459               Gaps:       19
 Percent Similarity:  72.705   Percent Identity: 50.496

alignment_block:
 US-09-634-252A-2 x Q9R1V4   ..

 Align seg 1/1  to: Q9R1V4  from: 1  to: 773

     139 CTTCTCGTCCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCG 188
         :::   :::|||||||||||||||   |||               |||
       9 IleAlaAlaLeuLeuLeuLeuProLeuLeuPro.............ProPr 21

     189 CGCCTGGGGGGCTGCTGCGCCC...AGCGCTCCGCATTGGAATGAAACTG 235
          :::   ||||||   :::|||   :::|||   ||||||:::::::::|
      21 oGlyLeuGlyAlaLeuGlyProArgGlyAlaLeuHisTrpArgSerSerA 38

     236 CAGAAAAAAATTTGGGAGTCCTGGCAGATGAAGACAATACATTGCAACAG 285
         ||      ::::::|||
      38 la......HisValGly................................ 41

     286 AATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAATCAC 335
                          |||   ::::::         :::|||:::||
      42 ..................SerProGluSerProGluGlySerGluValTh 52

     336 ACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGAAAGCCCTT 385
         |   ||||||||||||:::         ::::::   :::|||
      52 rGluProSerArgLeuValArgGln...SerSerGlyGlyGluValArgL 68

     386 ATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAGGCT 435
                 |||||||||::::::|||::::::      ::::::
      68 ysProGlnLeuAspThrArgValArgGlnAspProProArgGlyThrPro 84

     436 GTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATT 485
         ||||||||||||||:::||||||   |||   ||||||   |||:::||
      85 ValHisLeuAlaGlnValSerPheValIleProAlaPheAspSerAsnPh 101

     486 CATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGG 535
         |   ||||||||   ||||||:::   ||||||||||||:::|||||||
     101 eThrLeuAspLeuGluLeuAsnHisHisLeuLeuSerSerGlnTyrValG 118

     536 AGATTCACTAC...GAAAATGGGAAACCACAGTACTCTAAGGGT...GGA 579
         ||   |||:::   ::::::|||      |||:::|||   |||   |||
     118 luArgHisPheSerArgGluGlyThrArgGlnHisSerThrGlyAlaGly 134

     580 GAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGT 629
         :::|||||||||||||||||:::::||||||      :::|||   ::
     135 AspHisCysTyrTyrHisGlyLysLeuArgGlyAsnProGlnSerPheAl 151

     630 GGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATGATACCT 679
         :|||||||||||||||:::|||||||||||:::||:::|||   ::::
     151 aAlaLeuSerThrCysGlnGlyLeuHisGlyValPheSerAspGlyAsnL 168

     680 TCGTGTATATGATAGAGCCACTAGAGCTGGTT......CATGATGAGAAA 723
         :::::|||::::::||||||   |||::::::      :::
     168 euThrTyrIleValGluProLysGluIleAlaGlyProArgGlyProPro 184

     724 AGCACAGGTCGACCACATATAATCCAGAAAACC..................... 756
```

```
 185 GlnGlyProLeuProHisLeuIleTyrArgThrProLeuLeuProAlaPr 201

 757 ..........................TTGGCAGGACAGTATTCTAAGC 778
                                 :::|||         ::::::
 201 oLeuGlyCysArgGluProGlyCysLeuPheAlaValProAlaGlnSerT 218

 779 AAATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCT 828
       :::   |||                              ||||||   |||
 218 hrLeuProAsn.......................TrpProLysLeu... 225

 829 GAATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGG 878
                      ||||||||||||   |||:::      |||||
 226 ..................ArgArgLysArgGlnValArg......ArgGl 234

 879 T.........ATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTA 919
     |         :::   :::|||   ||||||:::||||||:::::::::|
 234 yHisProThrValHisSerGluThrLysTyrValGluLeuIleValIleA 251

 920 ATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAAC 969
     ||||||||:::   :::::::::   |||:::|||   :::   |||:::
 251 snAspHisGlnLeuPheGluGlnMetArgGlnSerValValLeuThrSer 267

 970 AACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCA 1019
     |||||||||||||||||||||||||||:::|||   ||||||||||||||
 268 AsnPheAlaLysSerValValAsnLeuAlaAspValIleTyrLysGluGl 284

1020 GCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGG 1069
     ||||||||||||||:::|||||||||||||:::||||||||:::::   |
 284 nLeuAsnThrArgIleValLeuValAlaMetGluThrTrpAlaAspGlyA 301

1070 ATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCA 1119
     ||:::|||:::::      :::   ::::::   |||   ::::::
 301 spLysIleGlnValGlnAspAspLeuLeuGluThrLeuAlaArgLeuMet 317

1120 AAATACCGGCAGCGC...ATTAAGCAGCATGCTGATGCTGTGCACCTCAT 1166
        ||||||:::::   :::   :::   :::||||||:::||||||::
 318 ValTyrArgArgGluGlyLeuProGluProSerAspAlaThrHisLeuPh 334

1167 CTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAG 1216
     :|||      ||||||            |||:::   :::|||   ||||
 334 eSerGlyArgThrPheLysSerThrSerSerGlyAlaAlaTyrValGlyG 351

1217 GTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCA 1266
     ||:::|||||   :::|||||   ||||||||||||||||
 351 lyIleCysSerLeuSerArgGlyGlyGlyValAsnGluTyrGlyAsnMet 367

1267 ATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAAT 1316
        |||:::|||   :::|||:::|||:::|||:::|||||||||||::
 368 GlyAlaMetAlaValThrLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMe 384

  17 CCAATGG...GAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAAT 1363
     ::::|||   :::      |||           |||   |||   :::
  84 tMetTrpAsnLysHisArgSerSerAlaGlyAspCysLysCysProAspI 401

1364 CCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAA 1413
       |||   |||||||||||||||:::||||||            ||||||
 401 leTrpLeuGlyCysIleMetGluAspThrGlyPheTyrLeuProArgLys 417

1414 TTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGG 1463
     ||||||:::|||||||||   ||||||:::::|||||||||:::||||||
 418 PheSerArgCysSerIleAspGluTyrAsnGlnPheLeuGlnGluGlyGl 434

1464 TGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAAT 1513
     ||||:::||||||||||||:::|||   ||||||:::::|||   ||||
 434 yGlySerCysLeuPheAsnLysProLeuLysLeuLeuAspProProGluC 451

1514 GTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCAT 1563
     ||||||||||:::|||||||||||||||||||||||||||||||||||
 451 ysGlyAsnGlyPheValGluAlaGlyGluGluCysAspCysGlySerVal 467

1564 GTGGAATGCTAT.........GGATTATGCTGTAAGAAATGTTCCCTCTC 1604
        ||||||            |||   |||||||||||||||:::|||::

 468 GlnGluCysSerArgAlaGlyGlyAsnCysCysLysLysCysThrLeuTh 484

1605 CAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTC 1654
     ::::   |||   ||||||||||||   |||||::::::      |||
 484 rHisAspAlaMetCysSerAspGlyLeuCysCysArgArg......CysL 499

1655 TTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGAT 1704
       ::::::|||||||||   :::|||||||:::|||||||||||||||||
 499 ysTyrGluProArgGlyValSerCysArgGluAlaValAsnGluCysAsp 515

1705 ATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCA 1754
     |||:::|||   |||||||||||||||:::|||||||||||||||||||
 516 IleAlaGluThrCysThrGlyAspSerSerGlnCysProProAsnLeuHi 532

1755 TAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATG 1804
     ||||   ||||||||   |||:::::::::|||||||||||||||:::|
 532 sLysLeuAspGlyTyrTyrCysAspHisGluGlnGlyArgCysTyrGlyG 549

1805 GCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAG 1854
     ||:::|||||||||||||||:::|||||||||   :::|||||||
 549 lyArgCysLysThrArgAspArgGlnCysGlnAlaLeuTrpGlyHisAla 565

1855 GCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCAC 1904
     ||||||      |||:::|||||||||||||||||||:::||||||||
 566 AlaAla......AspArgPheCysTyrGluLysLeuAsnValGluGlyTh 580

1905 TGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCA 1954
     ||||:::|||||||||||:::   |||:::   |||:::||||||||||
 580 rGluArgGlyAsnCysGlyArgLysGlySerGlyTrpValGlnCysSerL 597

1955 AACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCT 2004
     ||:::||||||:::|||||||||||||||||:::|||:::::   |||
 597 ysGlnAspValLeuCysGlyPheLeuLeuCysValAsnIleSerGlyAla 613

2005 CCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCA 2054
     ||||||:::|||:::|||   |||:::|||      ::::::||||||||
 614 ProArgLeuGlyAspLeuGlyGlyAspIleSerSerValThrPheTyrHi 630

2055 TCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATG 2104
     ||||||||:::   :::||||||   |||:::||||||   |||   |||
 630 sGlnGlyLysGluLeuAspCysArgGlyGlyHisValGlnLeuAlaAspG 647

2105 ATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATG 2154
       :::|||:::::|||||||||||||||||||   ||||||||||:::|||
 647 lySerAspLeuSerTyrValGluAspGlyThrAlaCysGlyProAsnMet 663

2155 ATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAG 2204
     :::|||||||||:::::::|||||      :::|||:::|||:::|||::
 664 LeuCysLeuAspHisArgCysLeuProAlaSerAlaPheAsnPheSerTh 680

2205 CTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTA 2254
     :||||||   :::::::::   ::::::||||||   ||||||||||||
 680 rCysProGlySerGlyGluArgArgIleCysSerHisHisGlyValCysS 697

2255 GTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGC 2304
     |||||||||::   ||||||||||::      |||:::|||   ||||||
 697 erAsnGluGlyLysCysIleCysGlnProAspTrpThrGlyLysAspCys 713

2305 AGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCC...AAGGATGAAGG 2351
     |||||::::::|||:::   :::      |||      :::   |||
 714 SerIleHisAsnProLeuProThrSerProProThrGlyGluThrGluAr 730

2352 ACCCAAGGGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGGTG 2401
        ||||||||||||:::|||||:::|||||||||||||||||||||||
 730 gTyrLysGlyProSerGlyThrAsnIleIleIleGlySerIleAlaGlyA 747

2402 CCATCCTGGTAGCAGCTATTGTCCTTGGGGGCACAGGCTGGGGATTTAAA 2451
     ||:::|||||||||||||||||||||||||||||||||||||||||||||
 747 laValLeuValAlaAlaIleValLeuGlyGlyThrGlyTrpGlyPheLys 763

2452 AATGTCAAGAAGAGAAGG 2469
     |||:::::::::   |||
 764 AsnIleArgArgGlyArg 769
```

```
seq_name: sp_human:Q9UIA1

seq_documentation_block:
ID   Q9UIA1      PRELIMINARY;      PRT;   870 AA.
AC   Q9UIA1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MDC2.
GN   MDC2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
     NCBI_TaxID=9606;
     [1]
     SEQUENCE FROM N.A.
     TISSUE=BRAIN;
     Wada M., Torigoe K., Harada T., Kuwano M.;
RT   "Isolation and tissue specific expression of novel ADAM family from
RT   7q21.1 region.";
RL   Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF073291; AAF22476.2; -.
DR   HSSP; P18619; 1FVL.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PRINTS; PR00289; DISINTEGRIN.
DR   ProDom; PD000664; -; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   SMART; SM00050; DISIN; 1.
SQ   SEQUENCE   870 AA;  96695 MW;  3BDF8E14213F860E CRC64;


alignment_scores:
            Quality: 1985.50              Length:    817
              Ratio:  3.429                Gaps:     14
 Percent Similarity:  70.869   Percent Identity: 47.491

alignment_block:
 09-634-252A-2 x Q9UIA1   ..

gn seg 1/1  to: Q9UIA1  from: 1  to: 870

    106 AGCGCCCCGGCCCGCACGCCGCCCTGCCGCCTGCTTCTCGTCCTTCTCCT 155
        :::|||   |||   :::   |||       ||||||   ||||||
      3 AlaAlaValAlaValSerValProPheLeuLeuLeuCysValLeuGlyTh 19

    156 GCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCGCGCCTGGGGGGCTGCTG 205
            ||||||      :::                        |||   || :
     19 rCysProProAlaArgCys.....................GlyGlnAlaG 29

    206 CGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAAAAAAATTTGGGAGTC 255
        ::   ::::::            |||      :::::::|||
     29 lyAspAlaSerLeuMetGluLeuGluLysArgLysGluAsn......... 42

    256 CTGGCAGATGAAGACAATACATTGCAACAGAATAGCAGCAGTAATATCAG 305

     42 ............................................... 42

    306 TTACAGCAATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATAT 355
              :::   ::::::::::::      :::|||   ||||||||||
     43 .......ArgPheValGluArgGlnSerIleValProLeuArgLeuIleT 57

    356 ATTACATCAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAG 405
        ||      :::   ::::::||||||   :::   :::||||||||||:::
     57 yrArgSerGlyGlyGluAspGluSerArgHisAspAlaLeuAspThrArg 73

    406 GCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAG 455
        :::|||   :::         ::::::   :::|||:::   ||||||||
     74 ValArgGlyAspLeuGlyGlyArgGlnLeuThrHisValAspGlnAlaSe 90

    456 CTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGA 505
        |||||||::::::|||||||||:::::|||||||||||||:::::||||
     90 rPheGlnValAspAlaPheGlyThrSerPheIleLeuAspValValLeuA 107

    506 ACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAAAT... 552
        ||:::   ||||||||||||:::|||:::|||   |||   |||:::
    107 snHisAspLeuLeuSerSerGluTyrIleGluArgHisIleGluHisGly 123

    553 GGGAAACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAG 602
        ||||||            |||||||||||||||||||||||:::|||
    124 GlyLysThrValGluValLysGlyGlyGluHisCysTyrTyrGlnGlyHi 140

    603 CATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGAC 652
         |||||||||      ||||||   ||||||||||||||||||:::||||
    140 sIleArgGlyAsnProAspSerPheValAlaLeuSerThrCysHisGlyL 157

    653 TTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTA 702
        ||||||||||||||   |||   :::   :::|||:::|||||||||
    157 euHisGlyMetPheTyrAspGlyAsnHisThrTyrLeuIleGluPro... 172

    703 GAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCA.............CA 740
                   :::|||:::::::|||   :::                ||
    173 .............GluGluAsnAspThrThrGlnGluAspPheHisPheHi 185

    741 TATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAG.............. 777
        |   :::   |||:::          ::::::|||
    185 sSerValTyrLysSerArgLeuPheGluPheSerLeuAspAspLeuProS 202

    778 ........CAAATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCC 819
                ||||:::   |||:::|||   ::::::
    202 erGluPheGlnGlnValAsnIleThrProSerLys.............. 213

    820 TTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCC 869
        |||:::   :::   :::       :::   :::::::::      :::
    214 PheIleLeuLysProArgProLysArgSerLysArgGlnLeuArgArgTy 230

    870 ATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTA 919
           |||::::::   ||||||   ||||||:::||||||||||||||||
    230 rProArgAsnValGluGluGluThrLysTyrIleGluLeuMetIleValA 247

    920 ATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAAC 969
        ||||||||      :::||||||||||||   |||   :::||||||||||
    247 snAspHisLeuMetPheLysLysHisArgLeuSerValValHisThrAsn 263

    970 AACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCA 1019
        ::::::|||||||||||||||||||:::::|||   ||||||||||:::||
    264 ThrTyrAlaLysSerValValAsnMetAlaAspLeuIleTyrLysAspGl 280

   1020 GCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGG 1069
        ||||:::||||||:::|||||||||||:::||||||||:::      :
    280 nLeuLysThrArgIleValLeuValAlaMetGluThrTrpAlaThrAspA 297

   1070 ATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCA 1119
        ::::::::   |||:::   ||||||:::      |||:::||||||
    297 snLysPheAlaIleSerGluAsnProLeuIleThrLeuArgGluPheMet 313

   1120 AAATACCGGCAGCGC...ATTAAGCAGCATGCTGATGCTGTGCACCTCAT 1166
        ||||||||||:::      ||||||:::   :::|||||||||||||::
    314 LysTyrArgArgAspPheIleLysGluLysSerAspAlaValHisLeuPh 330

   1167 CTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAG 1216
        :|||         |||:::   :::||||||:::   :::|||:::||||
    330 eSerGlySerGlnPheGluSerSerArgSerGlyAlaAlaTyrIleGlyG 347

   1217 GTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCA 1266
        ||:::||||||      :::|||   ||||||||||||:::|||
    347 lyIleCysSerLeuLeuLysGlyGlyGlyValAsnGluPheGlyLysThr 363
```

```
1267 ATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAAT 1316
364 AspLeuMetAlaValThrLeuAlaGlnSerLeuAlaHisAsnIleGlyIl 380

1317 CCAATGGGAACCTTCTAGCAGAAAGCCAAAA.............TGTG 1351
380 e............IleSerAspLysArgLysLeuAlaSerGlyGluCysL 393

1352 ACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCC 1401
393 ysCysGluAspThrTrpSerGlyCysIleMetGlyAspThrGlyTyrTyr 409

1402 CATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTT 1451
410 LeuProLysLysPheThrGlnCysAsnIleGluGluTyrHisAspPheLe 426

1452 ACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTG 1501
426 uAsnSerGlyGlyGlyAlaCysLeuPheAsnLysProSerLysLeuLeuA 443

502 AGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGAT 1551
443 spProProGluCysGlyAsnGlyPheIleGluThrGlyGluGluCysAsp 459

1552 TGTGGTTTTCATGTGGAATGCTAT.........GGATTATGCTGTAAGAA 1592
460 CysGlyThrProAlaGluCysValLeuGluGlyAlaGluCysCysLysLy 476

1593 ATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACA 1642
476 sCysThrLeuThrGlnAspSerGlnCysSerAspGlyLeuCysCysLys. 492

1643 ATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTG 1692
493 .....LysCysLysPheGlnProMetGlyThrValCysArgGluAlaVal 507

1693 AACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCC 1742
508 AsnAspCysAspIleArgGluThrCysSerGlyAsnSerSerGlnCysAl 524

1743 ACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCC 1792
524 aProAsnIleHisLysMetAspGlyTyrSerCysAspGlyValGlnGlyI 541

1793 GCTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATC 1842
541 leCysPheGlyGlyArgCysLysThrArgAspArgGlnCysLysTyrIle 557

1843 TGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAA 1892
558 TrpGlyGlnLysValThrAlaSerAspLysTyrCysTyrGluLysLeuAs 574

1893 TACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGA 1942
574 nIleGluGlyThrGluLysGlyAsnCysGlyLysAspLysAspThrTrpI 591

1943 TTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAAT 1992
591 leGlnCysAsnLysArgAspValLeuCysGlyTyrLeuLeuCysThrAsn 607

1993 CTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAAC 2042
608 IleGlyAsnIleProArgLeuGlyGluLeuAspGlyGluIleThrSerTh 624

2043 TTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAG 2092
624 rLeuValValGlnGlnGlyArgThrLeuAsnCysSerGlyGlyHisValL 641

2093 TTTTAGATGATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGT 2142
641 ysLeuGluGluAspValAspLeuGlyTyrValGluAspGlyThrProCys 657

2143 GGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCT 2192

658 GlyProGlnMetMetCysLeuGluHisArgCysLeuProValAlaSerPh 674

2193 AAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCC 2242
674 eAsnPheSerThrCysLeuSerSerLysGluGlyThrIleCysSerGlyA 691

2243 ATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCA 2292
691 snGlyValCysSerAsnGluLeuLysCysValCysAsnArgHisTrpIle 707

2293 GGGACAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAA 2342
708 GlySerAspCys.....AsnThrTyrPheProHi 717

2343 GGATGAAGGACCCAAG.....GGTCCTAGTG 2368
717 sAsnAspAspAlaLysThrGlyIleThrLeuSerGlyAsnGlyValAlaG 734

2369 CCACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCT 2418
734 lyThrAsnIleIleIleGlyIleIleAlaGlyThrIleLeuValLeuAla 750

2419 ATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAAATGTCAAGAAGAGAAG 2468
751 LeuIleLeuGlyIleThrAlaTrpGlyTyrLysAsnTyrArgGluGlnAr 767

2469 G 2469
767 g 767
```

seq_name: sp_human:Q9UKK2

seq_documentation_block:

```
ID   Q9UKK2      PRELIMINARY;      PRT;   867 AA.
AC   Q9UKK2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   METALLOPROTEINASE-DISINTEGRIN ADAM22-3 (FRAGMENT).
GN   ADAM22.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CEREBELLUM;
RX   MEDLINE=99453762; PubMed=10524237;
RA   Poindexter K., Nelson N., DuBose R.F., Black R.A., Cerretti D.P.;
RT   "Identification of seven metalloproteinase-disintegrin (ADAM) genes
RT   from genomic libraries.";
RL   Gene 237:61-70(1999).
DR   EMBL; AF158637; AAD55251.1; -.
DR   HSSP; P18619; 1FVL.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PRINTS; PR00289; DISINTEGRIN.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   SMART; SM00050; DISIN; 1.
KW   Integrin.
FT   NON_TER       1      1
SQ   SEQUENCE   867 AA;  96462 MW;  974F668D893E7BDE CRC64;
```

alignment_scores:

```
          Quality: 1985.00              Length:    746
            Ratio:  3.564                Gaps:     12
Percent Similarity:  74.665   Percent Identity: 50.268

alignment_block:
US-09-634-252A-2 x Q9UKK2   ..

Align seg 1/1  to: Q9UKK2  from: 1  to: 867

    319 ATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACATCAACCA 368
          ::::::::::::       :::|||   ||||||||||||      :::
      6 ValGluArgGlnSerIleValProLeuArgLeuIleTyrArgSerGlyGl 22

    369 AGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGC 418
         ::::::||||||   :::   :::||||||||||:::::::|||   :::
     22 yGluAspGluSerArgHisAspAlaLeuAspThrArgValArgGlyAspL 39

    419 AAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAA 468
              ::::::   :::|||:::   |||||||||||||||:::::::
     39 euGlyGlyArgGlnLeuThrHisValAspGlnAlaSerPheGlnValAsp 55

    469 GCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGTT 518
         |||||||||:::::|||||||||||:::::|||||||:::   |||||
     56 AlaPheGlyThrSerPheIleLeuAspValValLeuAsnHisAspLeuLe 72

    519 GTCTTCTGATTATGTGGAGATTCACTACGAAAAT...GGGAAACCACAGT 565
         |||||||:::|||:::|||   |||   |||:::   ||||||
     72 uSerSerGluTyrIleGluArgHisIleGluHisGlyGlyLysThrValG 39

    566 ACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGCGTC 615
              |||||||||||||||||||||||:::|||   |||||||||
     89 luValLysGlyGlyGluHisCysTyrTyrGlnGlyHisIleArgGlyAsn 105

    616 AAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTT 665
         ||||||   ||||||||||||||||||:::|||||||||||||||||
     106 ProAspSerPheValAlaLeuSerThrCysHisGlyLeuHisGlyMetPh 122

    666 TGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATG 715
         |   |||   :::   :::|||:::||||||||                :
     122 eTyrAspGlyAsnHisThrTyrLeuIleGluPro.............G 134

    716 ATGAGAAAAGCACAGGTCGACCA............CATATAATCCAGAAA 753
         ::|||::::::|||   :::               |||   :::   |||
     134 luGluAsnAspThrThrGlnGluAspPheHisPheHisSerValTyrLys 150

    754 ACCTTGGCAGGACAGTATTCTAAG....................CAAAT 782
         :::         ::::::|||                      |||::
     151 SerArgLeuPheGluPheSerLeuAspAspLeuProSerGluPheGlnGl 167

    783 GAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAAT 832
         :   |||:::|||   ::::::               |||:::   :::
     167 nValAsnIleThrProSerLys..............PheIleLeuLysP 179

    833 TACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGTATA 882
           :::      :::   :::::::::      :::      |||::::::
     179 roArgProLysArgSerLysArgGlnLeuArgArgTyrProArgAsnVal 195

    883 TTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAAC 932
         ||||||   ||||||:::||||||||||||||||||||||||
     196 GluGluGluThrLysTyrIleGluLeuMetIleValAsnAspHisLeuMe 212

    933 GTATAAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGT 982
         :::||||||||||||   |||   :::|||||||||:::::|||||
     212 tPheLysLysHisArgLeuSerValValHisThrAsnThrTyrAlaLysS 229

    983 CCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGG 1032
         ||||||||||:::::|||   |||||||||:::||||||:::||||||
     229 erValValAsnMetAlaAspLeuIleTyrLysAspGlnLeuLysThrArg 245

   1033 GTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACAT 1082
         :::||||||||||||:::|||||||||:::      :::::::::   ||
     246 IleValLeuValAlaMetGluThrTrpAlaThrAspAsnLysPheAlaIl 262

   1083 CACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGC 1132
         |:::   ||||||:::      |||:::||||||   ||||||||:::
      262 eSerGluAsnProLeuIleThrLeuArgGluPheMetLysTyrArgArgA 279

   1133 GC...ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACA 1179
              ||||||:::   :::|||||||||||||||:::|||
      279 spPheIleLysGluLysSerAspAlaValHisLeuPheSerGlySerGln 295

   1180 TTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCG 1229
         |||:::   :::||||||:::   :::|||:::||||||:::||||||
      296 PheGluSerSerArgSerGlyAlaAlaTyrIleGlyGlyIleCysSerLe 312

   1230 CACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCAC 1279
           :::|||   ||||||||||||:::|||            :::|||
      312 uLeuLysGlyGlyGlyValAsnGluPheGlyLysThrAspLeuMetAlaV 329

   1280 AAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCT 1329
          :::|||:::||||||||||||:::|||:::||||||
      329 alThrLeuAlaGlnSerLeuAlaHisAsnIleGlyIle..... 341

   1330 TCTAGCAGAAAGCCAAAA.............TGTGACTGCACAGAATC 1364
          |||   |||   |||                |||   |||   :::::
      342 IleSerAspLysArgLysLeuAlaSerGlyGluCysLysCysGluAspTh 358

   1365 CTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAAT 1414
         :|||:::||||||||||||   :::||||||              :::||||
      358 rTrpSerGlyCysIleMetGlyAspThrGlyTyrTyrLeuProLysLysP 375

   1415 TTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGT 1464
         ||::::::|||:::|||   ||||||:::|||||||||:::   ||||||
      375 heThrGlnCysAsnIleGluGluTyrHisAspPheLeuAsnSerGlyGly 391

   1465 GGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATG 1514
         ||||||||||||||||||:::|||:::||||||:::::|||   |||||
      392 GlyAlaCysLeuPheAsnLysProSerLysLeuLeuAspProProGluCy 408

   1515 TGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATG 1564
         ||||||||||:::::|||:::||||||||||||||||||||||     :
      408 sGlyAsnGlyPheIleGluThrGlyGluGluCysAspCysGlyThrProA 425

   1565 TGGAATGCTAT.........GGATTATGCTGTAAGAAATGTTCCCTCTCC 1605
         ::||||||            :::   |||||||||||||||:::|||:::
      425 laGluCysValLeuGluGlyAlaGluCysCysLysLysCysThrLeuThr 441

   1606 AACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCT 1655
         :::   ::::::||||||||||||   ||||||:::      :::|||
      442 GlnAspSerGlnCysSerAspGlyLeuCysCysLys......LysCysLy 456

   1656 TTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATA 1705
         |||||||||   |||       ||||||:::|||||||||:::||||||||
      456 sPheGlnProMetGlyThrValCysArgGluAlaValAsnAspCysAspI 473

   1706 TTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCAT 1755
         ||   |||   |||:::|||:::|||:::||||||   ||||||:::|||
      473 leArgGluThrCysSerGlyAsnSerSerGlnCysAlaProAsnIleHis 489

   1756 AAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGG 1805
         |||:::|||||||||:::|||:::      ||||||   |||:::::::||
      490 LysMetAspGlyTyrSerCysAspGlyValGlnGlyIleCysPheGlyGl 506

   1806 CGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGG 1855
         |:::|||||||||||||:::||||||:::||||||||||||   |||:
      506 yArgCysLysThrArgAspArgGlnCysLysTyrIleTrpGlyGlnLysV 523

   1856 CTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACT 1905
         ::::::::|||||||||:::|||||||||||||||||   |||||||||
      523 alThrAlaSerAspLysTyrCysTyrGluLysLeuAsnIleGluGlyThr 539

   1906 GAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAA 1955
         ||||||||||||||||||||||   |||   ||||||||||:::||
      540 GluLysGlyAsnCysGlyLysAspLysAspThrTrpIleGlnCysAsnLy 556

   1956 ACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTC 2005
```

```
     |:::||||||:::||||||:::||||||||||||||||:::   :::   |
 556 sArgAspValLeuCysGlyTyrLeuLeuCysThrAsnIleGlyAsnIleP 573

2006 CACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCAT 2055
     |||||:::|||:::|||:::|||||||||      |||         :::
 573 roArgLeuGlyGluLeuAspGlyGluIleThrSerThrLeuValValGln 589

2056 CAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGA 2105
     |||||||||:::::::::|||||||||:::||||||   |||:::::::||
 590 GlnGlyArgThrLeuAsnCysSerGlyGlyHisValLysLeuGluGluAs 606

2106 TACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGA 2155
     |:::|||:::|||||||||||||||||||||||||||||||||:::||||
 606 pValAspLeuGlyTyrValGluAspGlyThrProCysGlyProGlnMetM 623

2156 TGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGC 2205
     ||||||||:::::::::||||||   :::   ::::::|||:::|||:::
 623 etCysLeuGluHisArgCysLeuProValAlaSerPheAsnPheSerThr 639

2206 TGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAG 2255
     |||      ::::::::::|||   :::|||||||||:::||||||||||
 640 CysLeuSerSerLysGluGlyThrIleCysSerGlyAsnGlyValCysSe 656

2256 TAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCA 2305
     |||||||      |||:::|||:::      |||   |||:::||||||
 656 rAsnGluLeuLysCysValCysAsnArgHisTrpIleGlySerAspCys. 672

2306 GTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAAGGATGAAGGACCC 2355
                        |||   :::   |||   ::::::
 673 ...................AsnThrTyrPheProHisAsnAspAspAla 682

2356 AAG.....................GGTCCTAGTGCCACCAATCTCAT 2381
     |||                    |||   ::::::||||||:::||
 683 LysThrGlyIleThrLeuSerGlyAsnGlyValAlaGlyThrAsnIleIl 699

2382 AATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCTATTGTCCTTGGGG 2431
     |||||||   |||||||||:::|||||||||   |||:::::::||||||
 699 eIleGlyIleIleAlaGlyThrIleLeuValLeuAlaLeuIleLeuGlyI 716

2432 GCACAGGCTGGGGATTTAAAAATGTCAAGAAGAGAAGG 2469
       |||:::|||||||:::||||||   :::::::::|||
 716 leThrAlaTrpGlyTyrLysAsnTyrArgGluGlnArg 728
```

```
seq_name: sp_human:O75076

seq_documentation_block:
ID   O75076      PRELIMINARY;      PRT;   823 AA.
AC   O75076;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MDC2 BETA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=98359734; PubMed=9693107;
RA   Sagane K., Ohya Y., Hasegawa Y., Tanaka I.;
RT   "Metalloproteinase-like, disintegrin-like, cysteine-rich proteins MDC2
RT   and MDC3: novel human cellular disintegrins highly expressed in the
RT   brain.";
RL   Biochem. J. 334:93-98(1998).
DR   EMBL; AB009671; BAA32350.1; -.
DR   HSSP; P18619; 1FVL.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   ProDom; PD000664; -; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   SMART; SM00050; DISIN; 1.
SQ   SEQUENCE   823 AA;  91491 MW;  E9FD1A9BA99DE0DB CRC64;


alignment_scores:
            Quality: 1983.50              Length:    817
              Ratio:  3.432                 Gaps:     14
 Percent Similarity:  70.747   Percent Identity: 47.491

alignment_block:
 US-09-634-252A-2 x O75076   ..

 Align seg 1/1  to: O75076  from: 1  to: 823

    106 AGCGCCCCGGCCCGCACGCCGCCCTGCCGCCTGCTTCTCGTCCTTCTCCT 155
        :::|||   |||   :::   |||      ||||||   ||||||
      3 AlaAlaValAlaValSerValProPheLeuLeuLeuCysValLeuGlyTh 19

    156 GCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCGCGCCTGGGGGGCTGCTG 205
           ||||||      :::                  |||   |||:
     19 rCysProProAlaArgCys.....................GlyGlnAlaG 29

    206 CGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAAAAAAATTTGGGAGTC 255
        ::   :::::           |||      ::::::|||
     29 lyAspAlaSerLeuMetGluLeuGluLysArgLysGluAsn......... 42

    256 CTGGCAGATGAAGACAATACATTGCAACAGAATAGCAGCAGTAATATCAG 305

     42 ..................................................... 42

    306 TTACAGCAATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATAT 355
           :::   ::::::::::::      :::|||   ||||||||||
     43 .......ArgPheValGluArgGlnSerIleValProProArgLeuIleT 57

    356 ATTACATCAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAG 405
        ||      :::   ::::::||||||   :::   :::|||||||||:::
     57 yrArgSerGlyGlyGluAspGluSerArgHisAspAlaLeuAspThrArg 73

    406 GCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAG 455
        :::|||   :::            :::   :::|||:::   ||||||||
     74 ValArgGlyAspLeuGlyGlyProGlnLeuThrHisValAspGlnAlaSe 90

    456 CTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGA 505
        ||||||:::::|||||||:::::||||||||||||||:::::||||
     90 rPheGlnValAspAlaPheGlyThrSerPheIleLeuAspValValLeuA 107

    506 ACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAAAT... 552
        ||:::   ||||||||||||:::|||:::|||   |||   |||:::
    107 snHisAspLeuLeuSerSerGluTyrIleGluArgHisIleGluHisGly 123

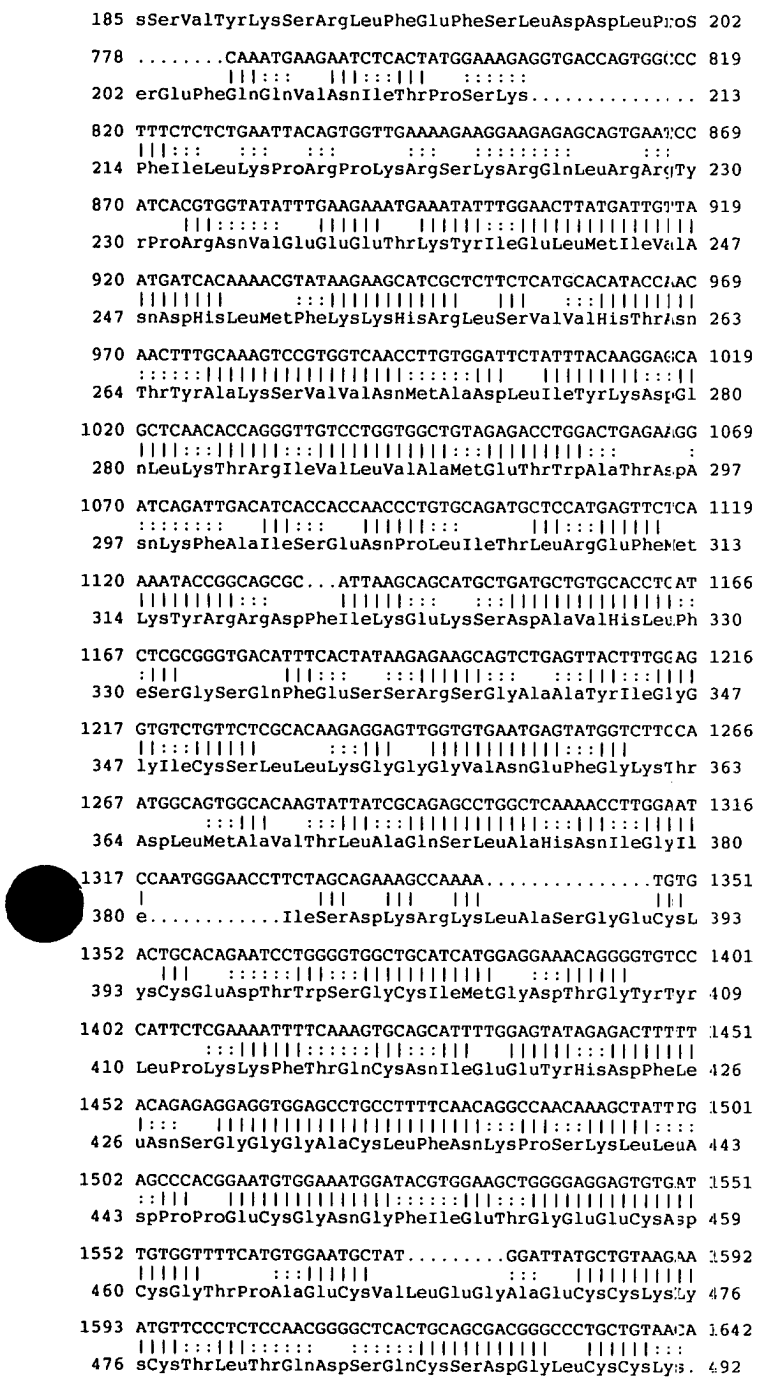
    553 GGGAAACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAG 602
        ||||||            |||||||||||||||||||||||:::|||
    124 GlyLysThrValGluValLysGlyGlyGluHisCysTyrTyrGlnGlyHi 140

    603 CATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGAC 652
         |||||||||      ||||||   ||||||||||||||||||:::||||
    140 sIleArgGlyAsnProAspSerPheValAlaLeuSerThrCysHisGlyL 157

    653 TTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTA 702
        ||||||||||||   |||   :::   :::|||:::|||||||||
    157 euHisGlyMetPheTyrAspGlyAsnHisThrTyrLeuIleGluPro... 172

    703 GAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCA.....CA 740
                   :::|||:::::::|||   :::              ||
    173 ............GluGluAsnAspThrThrGlnGluAspPheHisPheHi 185

    741 TATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAG..... 777
        |   :::   |||:::         ::::::|||
```


185 sSerValTyrLysSerArgLeuPheGluPheSerLeuAspAspLeuPro 202
778 CAAATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCC 819
202 erGluPheGlnGlnValAsnIleThrProSerLys..... 213
820 TTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAACCC 869
214 PheIleLeuLysProArgProLysArgSerLysArgGlnLeuArgArgTy 230
870 ATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTA 919
230 rProArgAsnValGluGluGluThrLysTyrIleGluLeuMetIleValA 247
920 ATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAAC 969
247 snAspHisLeuMetPheLysLysHisArgLeuSerValValHisThrAsn 263
970 AACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCA 1019
264 ThrTyrAlaLysSerValValAsnMetAlaAspLeuIleTyrLysAspGl 280
1020 GCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGG 1069
280 nLeuLysThrArgIleValLeuValAlaMetGluThrTrpAlaThrAspA 297
1070 ATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCA 1119
297 snLysPheAlaIleSerGluAsnProLeuIleThrLeuArgGluPheMet 313
1120 AAATACCGGCAGCGC...ATTAAGCAGCATGCTGATGCTGTGCACCTCAT 1166
314 LysTyrArgArgAspPheIleLysGluLysSerAspAlaValHisLeuPh 330
1167 CTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGCAG 1216
330 eSerGlySerGlnPheGluSerSerArgSerGlyAlaAlaTyrIleGlyG 347
1217 GTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCA 1266
347 lyIleCysSerLeuLeuLysGlyGlyValAsnGluPheGlyLysThr 363
1267 ATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAAT 1316
364 AspLeuMetAlaValThrLeuAlaGlnSerLeuAlaHisAsnIleGlyIl 380
1317 CCAATGGGAACCTTCTAGCAGAAAGCCAAAA.....TGTG 1351
380 e.....IleSerAspLysArgLysLeuAlaSerGlyGluCysL 393
1352 ACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCC 1401
393 ysCysGluAspThrTrpSerGlyCysIleMetGlyAspThrGlyTyrTyr 409
1402 CATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTT 1451
410 LeuProLysLysPheThrGlnCysAsnIleGluGluTyrHisAspPheLe 426
1452 ACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTG 1501
426 uAsnSerGlyGlyGlyAlaCysLeuPheAsnLysProSerLysLeuLeuA 443
1502 AGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGAT 1551
443 spProProGluCysGlyAsnGlyPheIleGluThrGlyGluGluCysAsp 459
1552 TGTGGTTTTCATGTGGAATGCTAT.....GGATTATGCTGTAAGAA 1592
460 CysGlyThrProAlaGluCysValLeuGluGlyAlaGluCysCysLysLy 476
1593 ATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACA 1642
476 sCysThrLeuThrGlnAspSerGlnCysSerAspGlyLeuCysCysLys. 492
1643 ATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTG 1692
493 LysCysLysPheGlnProMetGlyThrValCysArgGluAlaVal 507
1693 AACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCC 1742
508 AsnAspCysAspIleArgGluThrCysSerGlyAsnSerSerGlnCysAl 524
1743 ACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCC 1792
524 aProAsnIleHisLysMetAspGlyTyrSerCysAspGlyValGlnGlyI 541
1793 GCTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATC 1842
541 leCysPheGlyGlyArgCysLysThrArgAspArgGlnCysLysTyrIle 557
1843 TGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAA 1892
558 TrpGlyGlnLysValThrAlaSerAspLysTyrCysTyrGluLysLeuAs 574
1893 TACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGA 1942
574 nIleGluGlyThrGluLysGlyAsnCysGlyLysAspLysAspThrTrpI 591
1943 TTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAAT 1992
591 leGlnCysAsnLysArgAspValLeuCysGlyTyrLeuLeuCysThrAsn 607
1993 CTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAAC 2042
608 IleGlyAsnIleProArgLeuGlyGluLeuAspGlyGluIleThrSerTh 624
2043 TTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAG 2092
624 rLeuValValGlnGlnGlyArgThrLeuAsnCysSerGlyGlyHisValL 641
2093 TTTTAGATGATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGT 2142
641 ysLeuGluGluAspValAspLeuGlyTyrValGluAspGlyThrProCys 657
2143 GGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCT 2192
658 GlyProGlnMetMetCysLeuGluHisArgCysLeuProValAlaSerPh 674
2193 AAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCC 2242
674 eAsnPheSerThrCysLeuSerSerLysGluGlyThrIleCysSerGlyA 691
2243 ATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCA 2292
691 snGlyValCysSerAsnGluLeuLysCysValCysAsnArgHisTrpIle 707
2293 GGGACAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAA 2342
708 GlySerAspCys.....AsnThrTyrPheProHi 717
2343 GGATGAAGGACCCAAG.....GGTCCTAGTG 2368
717 sAsnAspAspAlaLysThrGlyIleThrLeuSerGlyAsnGlyValAlaG 734
2369 CCACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCT 2418
734 lyThrAsnIleIleIleGlyIleIleAlaGlyThrIleLeuValLeuAla 750
2419 ATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAAATGTCAAGAAGAGAAG 2468
751 LeuIleLeuGlyIleThrAlaTrpGlyTyrLysAsnTyrArgGluGlnAr 767
2469 G 2469
767 g 767

```
seq_name: sp_human:O75075

seq_documentation_block:
ID   O75075      PRELIMINARY;      PRT;   859 AA.
AC   O75075;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MDC2 ALPHA.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=98359734; PubMed=9693107;
RA   Sagane K., Ohya Y., Hasegawa Y., Tanaka I.;
RT   "Metalloproteinase-like, disintegrin-like, cysteine-rich proteins MDC2
     and MDC3: novel human cellular disintegrins highly expressed in the
     brain.";
     Biochem. J. 334:93-98(1998).
     EMBL; AB009671; BAA32349.1; -.
     HSSP; P18619; 1FVL.
DR   MEROPS; M12.978; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   ProDom; PD000664; -; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   SMART; SM00050; DISIN; 1.
SQ   SEQUENCE   859 AA;  95287 MW;  6105B681A1331D55 CRC64;


alignment_scores:
            Quality: 1983.50              Length:    817
              Ratio:  3.432                Gaps:     14
 Percent Similarity:  70.747   Percent Identity: 47.491

alignment_block:
 US-09-634-252A-2 x O75075   ..

 Align seg 1/1  to: O75075  from: 1  to: 859

    106 AGCGCCCCGGCCCGCACGCCGCCCTGCCGCCTGCTTCTCGTCCTTCTCCT 155
        :::|||   |||   :::   |||      ||||||   ||||||
      3 AlaAlaValAlaValSerValProPheLeuLeuLeuCysValLeuGlyTh 19

    156 GCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCGCGCCTGGGGGGCTGCTG 205
            ||||||     :::                  |||   |||:
     19 rCysProProAlaArgCys.....................GlyGlnAlaG 29

    206 CGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAAAAAAATTTGGGAGTC 255
        ::   ::::::           |||      :::::|||
     29 lyAspAlaSerLeuMetGluLeuGluLysArgLysGluAsn......... 42

    256 CTGGCAGATGAAGACAATACATTGCAACAGAATAGCAGCAGTAATATCAG 305

     42 ............................................... 42

    306 TTACAGCAATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATAT 355
               :::   ::::::::::::      :::|||   ||||||||||
     43 .......ArgPheValGluArgGlnSerIleValProLeuArgLeuIleT 57

    356 ATTACATCAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAG 405
        ||      :::   :::::||||||   :::   :::||||||||||:::
     57 yrArgSerGlyGlyGluAspGluSerArgHisAspAlaLeuAspThrArg 73

    406 GCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAG 455
        :::|||   :::            :::   :::|||:::   ||||||||
     74 ValArgGlyAspLeuGlyGlyProGlnLeuThrHisValAspGlnAlaSe 90

    456 CTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGA 505
        |||||||::::::||||||||:::::|||||||||||||:::::||||
     90 rPheGlnValAspAlaPheGlyThrSerPheIleLeuAspValValLeuA 107

    506 ACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAAAT... 552
        ||:::   ||||||||||||:::|||:::|||   |||   |||:::
    107 snHisAspLeuLeuSerSerGluTyrIleGluArgHisIleGluHisGly 123

    553 GGGAAACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAG 602
        ||||||            |||||||||||||||||||||:::|||
    124 GlyLysThrValGluValLysGlyGlyGluHisCysTyrTyrGlnGlyHi 140

    603 CATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGAC 652
         ||||||||      ||||||   ||||||||||||||||||:::||||
    140 sIleArgGlyAsnProAspSerPheValAlaLeuSerThrCysHisGlyL 157

    653 TTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTA 702
        ||||||||||||||   |||   :::   :::|||:::||||||||||
    157 euHisGlyMetPheTyrAspGlyAsnHisThrTyrLeuIleGluPro... 172

    703 GAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCA.............CA 740
                   :::|||::::::|||   :::                ||
    173 ............GluGluAsnAspThrThrGlnGluAspPheHisPheHi 185

    741 TATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAG.............. 777
        |   :::   |||:::         ::::::|||
    185 sSerValTyrLysSerArgLeuPheGluPheSerLeuAspAspLeuProS 202

    778 ........CAAATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCC 819
                |||:::   |||:::|||   ::::::
    202 erGluPheGlnGlnValAsnIleThrProSerLys.............. 213

    820 TTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCC 869
        |||:::   :::   :::      :::   :::::::::      :::
    214 PheIleLeuLysProArgProLysArgSerLysArgGlnLeuArgArgTy 230

    870 ATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTA 919
            |||::::::   ||||||   ||||||:::||||||||||||||||
    230 rProArgAsnValGluGluGluThrLysTyrIleGluLeuMetIleValA 247

    920 ATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAAC 969
        ||||||||      :::||||||||||||   |||   :::||||||||||
    247 snAspHisLeuMetPheLysLysHisArgLeuSerValValHisThrAsn 263

    970 AACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCA 1019
        ::::::||||||||||||||||||:::::|||   ||||||||||:::||
    264 ThrTyrAlaLysSerValValAsnMetAlaAspLeuIleTyrLysAspGl 280

   1020 GCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGG 1069
        ||||:::||||||:::||||||||||||:::||||||||||:::      :
    280 nLeuLysThrArgIleValLeuValAlaMetGluThrTrpAlaThrAspA 297

   1070 ATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCA 1119
        ::::::::   |||:::   ||||||:::      |||:::||||||
    297 snLysPheAlaIleSerGluAsnProLeuIleThrLeuArgGluPheMet 313

   1120 AAATACCGGCAGCGC...ATTAAGCAGCATGCTGATGCTGTGCACCTCAT 1166
        |||||||||:::      ||||||:::   :::||||||||||||||::
    314 LysTyrArgArgAspPheIleLysGluLysSerAspAlaValHisLeuPh 330

   1167 CTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAG 1216
        :|||         |||:::   :::|||||:::   :::|||:::||||
    330 eSerGlySerGlnPheGluSerSerArgSerGlyAlaAlaTyrIleGlyG 347

   1217 GTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCA 1266
        ||:::||||||      :::|||   ||||||||||||:::|||
    347 lyIleCysSerLeuLeuLysGlyGlyGlyValAsnGluPheGlyLysThr 363
```



seq_name: sp_human:Q9P0K2

seq_documentation_block:
ID Q9P0K2 PRELIMINARY; PRT; 899 AA.
AC Q9P0K2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE METALLOPROTEASE-LIKE, DISINTEGRIN-LIKE, CYSTEINE-RICH PROTEIN 2 DELTA.
GN ADAM22.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Harada T., Nishie A., Torigoe K., Ohtubo M., Ikezaki K., Shono T.,
RA Morioka T., Wada M., Kuwano M.;
RT "Identification of three novel splice variants of human
RT metalloprotease-like, disintegrin-like, cysteine-rich protein 2 gene
RT and their specific expression in brain tissue and gliomas.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF155381; AAF73288.1; -.
DR InterPro; IPR000561; -.
DR InterPro; IPR001590; -.
DR InterPro; IPR001762; -.
DR InterPro; IPR002870; -.
DR Pfam; PF00200; disintegrin; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR PRINTS; PR00289; DISINTEGRIN.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50214; DISINTEGRIN_2; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR SMART; SM00050; DISIN; 1.
KW Integrin; Protease.
SQ SEQUENCE 899 AA; 100340 MW; 5AB89C260378DFD6 CRC64;

alignment_scores:
Quality: 1983.50 Length: 817
Ratio: 3.432 Gaps: 14

Percent Similarity: 70.747 Percent Identity: 47.491

alignment_block:
US-09-634-252A-2 x Q9P0K2 ..

Align seg 1/1 to: Q9P0K2 from: 1 to: 899

```
106 AGCGCCCCGGCCCGCACGCCGCCCTGCCGCCTGCTTCTCGTCCTTCTCCT 155
    :::|||   |||   :::   |||      ||||||   ||||||
  3 AlaAlaValAlaValSerValProPheLeuLeuLeuCysValLeuGlyTh 19

156 GCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCGCGCCTGGGGGGCTGCTG 205
        ||||||      :::                  |||   |||:
 19 rCysProProAlaArgCys.....................GlyGlnAlaG 29

206 CGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAAAAAAATTTGGGAGTC 255
    ::   ::::::           |||      :::::::|||
 29 lyAspAlaSerLeuMetGluLeuGluLysArgLysGluAsn......... 42

256 CTGGCAGATGAAGACAATACATTGCAACAGAATAGCAGCAGTAATATCAG 305

 42 ................................................... 42

306 TTACAGCAATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATAT 355
        :::   ::::::::::::      :::|||   ||||||||||
 43 .......ArgPheValGluArgGlnSerIleValProLeuArgLeuIleT 57

356 ATTACATCAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAG 405
    ||      :::   :::::::|||||   :::   :::||||||||:::
 57 yrArgSerGlyGlyGluAspGluSerArgHisAspAlaLeuAspThrArg 73

406 GCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAG 455
    :::|||   :::            :::   :::|||:::   ||||||||
 74 ValArgGlyAspLeuGlyGlyProGlnLeuThrHisValAspGlnAlaSe 90

456 CTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGA 505
    |||||||::::::|||||||||::::::||||||||||::::::||||
 90 rPheGlnValAspAlaPheGlyThrSerPheIleLeuAspValValLeuA 107

506 ACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAAAT... 552
    ||:::   ||||||||||||:::|||:::|||   |||   |||:::
107 snHisAspLeuLeuSerSerGluTyrIleGluArgHisIleGluHisGly 123

553 GGGAAACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAG 602
    ||||||            ||||||||||||||||||||||:::|||
124 GlyLysThrValGluValLysGlyGlyGluHisCysTyrTyrGlnGlyHi 140

603 CATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGAC 652
     |||||||||      ||||||   |||||||||||||||||:::||||
140 sIleArgGlyAsnProAspSerPheValAlaLeuSerThrCysHisGlyL 157

653 TTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTA 702
    ||||||||||||||   |||   :::   :::|||:::||||||||||
157 euHisGlyMetPheTyrAspGlyAsnHisThrTyrLeuIleGluPro... 172

703 GAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCA............CA 740
                :::|||::::::|||   :::                ||
173 ............GluGluAsnAspThrThrGlnGluAspPheHisPheHi 185

741 TATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAG.............. 777
    |   :::   |||:::          :::::::|||
185 sSerValTyrLysSerArgLeuPheGluPheSerLeuAspAspLeuProS 202

778 ........CAAATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCC 819
            |||:::   |||:::|||   ::::::
202 erGluPheGlnGlnValAsnIleThrProSerLys.............. 213

820 TTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCC 869
    |||:::   :::   :::      :::   :::::::::      :::
214 PheIleLeuLysProArgProLysArgSerLysArgGlnLeuArgArgTy 230

870 ATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTA 919
        |||::::::   ||||||   |||||||:::||||||||||||||||
```

```
230 rProArgAsnValGluGluGluThrLysTyrIleGluLeuMetIleValA 247

920 ATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAAC 969
    ||||||||      :::||||||||||||   |||   :::|||||||||
247 snAspHisLeuMetPheLysLysHisArgLeuSerValValHisThrAsn 263

970 AACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCA 1019
    ::::::||||||||||||||||||:::::|||   ||||||||||:::||
264 ThrTyrAlaLysSerValValAsnMetAlaAspLeuIleTyrLysAspGl 280

1020 GCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGG 1069
     ||||:::||||||:::|||||||||||||:::|||||||||:::      :
280 nLeuLysThrArgIleValLeuValAlaMetGluThrTrpAlaThrAspA 297

1070 ATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCA 1119
     ::::::::   |||:::   ||||||:::      |||:::||||||
297 snLysPheAlaIleSerGluAsnProLeuIleThrLeuArgGluPheMet 313

1120 AAATACCGGCAGCGC...ATTAAGCAGCATGCTGATGCTGTGCACCTCAT 1166
     |||||||||:::      |||||||:::   :::||||||||||||||::
314 LysTyrArgArgAspPheIleLysGluLysSerAspAlaValHisLeuPh 330

1167 CTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAG 1216
     :|||         |||:::   :::||||||:::   :::|||:::||||
330 eSerGlySerGlnPheGluSerSerArgSerGlyAlaAlaTyrIleGlyG 347

1217 GTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCA 1266
     ||:::||||||      :::|||   ||||||||||||:::|||
347 lyIleCysSerLeuLeuLysGlyGlyGlyValAsnGluPheGlyLysThr 363

1267 ATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAAT 1316
           :::|||   :::|||:::||||||||||:::|||:::||||||
364 AspLeuMetAlaValThrLeuAlaGlnSerLeuAlaHisAsnIleGlyIl 380

1317 CCAATGGGAACCTTCTAGCAGAAAGCCAAAA.............TGTG 1351
     |                |||   |||   |||               |||
380 e............IleSerAspLysArgLysLeuAlaSerGlyGluCysL 393

1352 ACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCC 1401
       |||   ::::::|||:::||||||||||||   :::||||||
393 ysCysGluAspThrTrpSerGlyCysIleMetGlyAspThrGlyTyrTyr 409

1402 CATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTT 1451
           :::||||||:::::|||:::|||   ||||||:::||||||||
410 LeuProLysLysPheThrGlnCysAsnIleGluGluTyrHisAspPheLe 426

1452 ACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTG 1501
     |:::   |||||||||||||||||||||||:::|||:::||||||::::
426 uAsnSerGlyGlyGlyAlaCysLeuPheAsnLysProSerLysLeuLeuA 443

1502 AGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGAT 1551
     ::|||   |||||||||||||||:::::|||:::||||||||||||||||
443 spProProGluCysGlyAsnGlyPheIleGluThrGlyGluGluCysAsp 459

1552 TGTGGTTTTCATGTGGAATGCTAT.........GGATTATGCTGTAAGAA 1592
     ||||||      :::|||||||            :::   ||||||||||
460 CysGlyThrProAlaGluCysValLeuGluGlyAlaGluCysCysLysLy 476

1593 ATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACA 1642
     ||||:::|||::::::   ::::::||||||||||||   ||||||:::
476 sCysThrLeuThrGlnAspSerGlnCysSerAspGlyLeuCysCysLys. 492

1643 ATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTG 1692
          :::|||   |||||||||   |||      ||||||:::||||||
493 .....LysCysLysPheGlnProMetGlyThrValCysArgGluAlaVal 507

1693 AACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCC 1742
     |||:::||||||||   |||   |||:::|||:::|||:::||||||
508 AsnAspCysAspIleArgGluThrCysSerGlyAsnSerSerGlnCysAl 524

1743 ACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCC 1792
      ||||||:::||||||:::||||||||||:::|||:::      ||||||
524 aProAsnIleHisLysMetAspGlyTyrSerCysAspGlyValGlnGlyI 541
```

```
 1793 GCTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATC 1842
       |||::::::|||:::|||||||||||||||:::||||||:::||||||
  541 leCysPheGlyGlyArgCysLysThrArgAspArgGlnCysLysTyrIle 557

 1843 TGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAA 1892
       ||||||   |||::::::::::||||||||:::||||||||||||||||
  558 TrpGlyGlnLysValThrAlaSerAspLysTyrCysTyrGluLysLeuAs 574

 1893 TACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGA 1942
       |   ||||||||||||||||||||||||||||||   |||   ||||
  574 nIleGluGlyThrGluLysGlyAsnCysGlyLysAspLysAspThrTrpI 591

 1943 TTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAAT 1992
       ||||||||:::|||:::||||||:::||||||:::||||||||||||||
  591 leGlnCysAsnLysArgAspValLeuCysGlyTyrLeuLeuCysThrAsn 607

  993 CTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAAC 2042
       :::   :::   |||||||:::|||:::|||:::|||||||||     ||
  608 IleGlyAsnIleProArgLeuGlyGluLeuAspGlyGluIleThrSerTh 624

 2043 TTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAG 2092
       |         :::|||||||||:::::::::|||||||||:::||||||
  624 rLeuValValGlnGlnGlyArgThrLeuAsnCysSerGlyGlyHisValL 641

 2093 TTTTAGATGATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGT 2142
       |||::::::|||:::|||:::|||||||||||||||||||||||||||
  641 ysLeuGluGluAspValAspLeuGlyTyrValGluAspGlyThrProCys 657

 2143 GGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCT 2192
       ||||||:::||||||||||||:::::::::||||||   :::   :::::
  658 GlyProGlnMetMetCysLeuGluHisArgCysLeuProValAlaSerPh 674

 2193 AAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCC 2242
       :|||:::|||:::|||      :::::::::|||   :::||||||||:
  674 eAsnPheSerThrCysLeuSerSerLysGluGlyThrIleCysSerGlyA 691

 2243 ATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCA 2292
       ::|||||||||||||||||      |||:::|||:::      |||
  691 snGlyValCysSerAsnGluLeuLysCysValCysAsnArgHisTrpIle 707

 2293 GGGACAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAA 2342
       |||:::||||||                  |||   :::   |||
  708 GlySerAspCys.....................AsnThrTyrPheProHi 717

 2343 GGATGAAGGACCCAAG.....................GGTCCTAGTG 2368
       ::::::      |||                    |||   ::::
  717 sAsnAspAspAlaLysThrGlyIleThrLeuSerGlyAsnGlyValAlaG 734

 2369 CCACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCT 2418
       ::||||||:::|||||||||   |||||||||:::|||||||||   |||
  734 lyThrAsnIleIleIleGlyIleIleAlaGlyThrIleLeuValLeuAla 750

 2419 ATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAAATGTCAAGAAGAGAAG 2468
       ::::::||||||   |||:::||||||:::||||||   :::::::::||
  751 LeuIleLeuGlyIleThrAlaTrpGlyTyrLysAsnTyrArgGluGlnAr 767

 2469 G 2469
       |
  767 g 767
```

seq_name: sp_human:Q9P0K1

seq_documentation_block:

```
ID   Q9P0K1      PRELIMINARY;      PRT;   906 AA.
AC   Q9P0K1;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   METALLOPROTEASE-LIKE, DISINTEGRIN-LIKE, CYSTEINE-RICH PROTEIN 2
DE   EPSILON.
GN   ADAM22.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Harada T., Nishie A., Torigoe K., Ohtubo M., Ikezaki K., Shono T.,
RA   Morioka T., Wada M., Kuwano M.;
RT   "Identification of three novel splice variants of human
RT   metalloprotease-like, disintegrin-like, cysteine-rich protein 2 gene
RT   and their specific expression in brain tissue and gliomas.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF155382; AAF73289.1; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PRINTS; PR00289; DISINTEGRIN.
DR   ProDom; PD000664; -; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   SMART; SM00050; DISIN; 1.
KW   Integrin; Protease.
SQ   SEQUENCE   906 AA;  100432 MW;  265ECCD0FA6C088B CRC64;
```

alignment_scores:

Quality: 1983.50 Length: 817
Ratio: 3.432 Gaps: 14
Percent Similarity: 70.747 Percent Identity: 47.491

alignment_block:

US-09-634-252A-2 x Q9P0K1 ..

Align seg 1/1 to: Q9P0K1 from: 1 to: 906

```
  106 AGCGCCCCGGCCCGCACGCCGCCCTGCCGCCTGCTTCTCGTCCTTCTCCT 155
       :::|||   |||   :::   |||      ||||||   ||||||
    3 AlaAlaValAlaValSerValProPheLeuLeuLeuCysValLeuGlyTh 19

  156 GCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCGCGCCTGGGGGGCTGCTG 205
          ||||||      :::                        |||   |||:
   19 rCysProProAlaArgCys.....................GlyGlnAlaG 29

  206 CGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAAAAAAATTTGGGAGTC 255
       ::   ::::::              |||      ::::::|||
   29 lyAspAlaSerLeuMetGluLeuGluLysArgLysGluAsn......... 42

  256 CTGGCAGATGAAGACAATACATTGCAACAGAATAGCAGCAGTAATATCAG 305

   42 ..................................................... 42

  306 TTACAGCAATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATAT 355
             :::   ::::::::::::      :::|||   ||||||||||
   43 .......ArgPheValGluArgGlnSerIleValProLeuArgLeuIleT 57

  356 ATTACATCAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAG 405
       ||      :::   ::::::||||||   :::   :::||||||||:::
   57 yrArgSerGlyGlyGluAspGluSerArgHisAspAlaLeuAspThrArg 73

  406 GCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAG 455
       :::|||   :::            :::   :::|||:::   ||||||||
   74 ValArgGlyAspLeuGlyGlyProGlnLeuThrHisValAspGlnAlaSe 90

  456 CTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGA 505
       |||||||::::::||||||||:::::|||||||||||:::::||||
   90 rPheGlnValAspAlaPheGlyThrSerPheIleLeuAspValValLeuA 107

  506 ACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAAAT... 552
       ||:::   ||||||||||||:::|||:::|||   |||   |||:::
```

```
 107 snHisAspLeuLeuSerSerGluTyrIleGluArgHisIleGluHisGly 123

 553 GGGAAACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAG 602
     ||||||           ||||||||||||||||||||||||:::|||
 124 GlyLysThrValGluValLysGlyGlyGluHisCysTyrTyrGlnGlyHi 140

 603 CATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGAC 652
      |||||||||      ||||||   ||||||||||||||||||:::||||
 140 sIleArgGlyAsnProAspSerPheValAlaLeuSerThrCysHisGlyL 157

 653 TTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTA 702
     ||||||||||||||   |||   :::   :::|||:::|||||||||
 157 euHisGlyMetPheTyrAspGlyAsnHisThrTyrLeuIleGluPro... 172

 703 GAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCA............:CA 740
               :::|||::::::|||   :::                  ||
 173 ............GluGluAsnAspThrThrGlnGluAspPheHisPheHi 185

 741 TATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAG..............  777
     |   :::   |||:::         :::::::|||
 185 sSerValTyrLysSerArgLeuPheGluPheSerLeuAspAspLeuProS 202

 778 ........CAAATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCC 819
             |||:::   |||:::|||   ::::::
 202 erGluPheGlnGlnValAsnIleThrProSerLys.............. 213

 820 TTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCC 869
     |||:::   :::   :::      :::   :::::::::      :::
 214 PheIleLeuLysProArgProLysArgSerLysArgGlnLeuArgArgTy 230

 870 ATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTA 919
         |||::::::   ||||||   ||||||:::||||||||||||||||
 230 rProArgAsnValGluGluGluThrLysTyrIleGluLeuMetIleValA 247

 920 ATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAAC 969
     ||||||||      :::||||||||||||   |||   :::|||||||||
 247 snAspHisLeuMetPheLysLysHisArgLeuSerValValHisThrAsn 263

 970 AACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCA 1019
     ::::::||||||||||||||||||::::::|||   ||||||||||:::||
 264 ThrTyrAlaLysSerValValAsnMetAlaAspLeuIleTyrLysAspGl 280

1020 GCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGG 1069
     ||||:::||||||:::||||||||||||:::|||||||||:::      :
 280 nLeuLysThrArgIleValLeuValAlaMetGluThrTrpAlaThrAspA 297

1070 ATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCA 1119
     ::::::::   |||:::   ||||||:::      |||:::||||||
 297 snLysPheAlaIleSerGluAsnProLeuIleThrLeuArgGluPheMet 313

1120 AAATACCGGCAGCGC...ATTAAGCAGCATGCTGATGCTGTGCACCTCAT 1166
     |||||||||:::      ||||||:::   :::||||||||||||||::
 314 LysTyrArgArgAspPheIleLysGluLysSerAspAlaValHisLeuPh 330

1167 CTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAG 1216
     :|||         |||:::   :::||||||:::   :::|||:::||||
 330 eSerGlySerGlnPheGluSerSerArgSerGlyAlaAlaTyrIleGlyG 347

1217 GTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCA 1266
     ||:::||||||      :::|||   ||||||||||||:::|||
 347 lyIleCysSerLeuLeuLysGlyGlyGlyValAsnGluPheGlyLysThr 363

1267 ATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAAT 1316
          :::|||   :::|||:::||||||||||||:::|||:::||||||
 364 AspLeuMetAlaValThrLeuAlaGlnSerLeuAlaHisAsnIleGlyIl 380

1317 CCAATGGGAACCTTCTAGCAGAAAGCCAAAA.............TGTG 1351
     |              |||   |||   |||               |||
 380 e............IleSerAspLysArgLysLeuAlaSerGlyGluCysL 393

1352 ACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCC 1401
       |||   ::::::|||:::||||||||||||   :::||||||
 393 ysCysGluAspThrTrpSerGlyCysIleMetGlyAspThrGlyTyrTyr 409
```

```
1402 CATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTT 1451
           :::||||||:::::::|||:::|||   ||||||:::||||||||
  410 LeuProLysLysPheThrGlnCysAsnIleGluGluTyrHisAspPheLe 426

1452 ACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTG 1501
     |:::   |||||||||||||||||||||||:::|||:::||||||:::::
  426 uAsnSerGlyGlyGlyAlaCysLeuPheAsnLysProSerLysLeuLeuA 443

1502 AGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGAT 1551
     ::|||   ||||||||||||||:::::::|||:::||||||||||||||
  443 spProProGluCysGlyAsnGlyPheIleGluThrGlyGluGluCysAsp 459

1552 TGTGGTTTTCATGTGGAATGCTAT.........GGATTATGCTGTAAGAA 1592
     ||||||       :::||||||             :::   ||||||||||
  460 CysGlyThrProAlaGluCysValLeuGluGlyAlaGluCysCysLysLy 476

1593 ATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACA 1642
     ||||:::|||::::::   ::::::||||||||||||   ||||||:::
  476 sCysThrLeuThrGlnAspSerGlnCysSerAspGlyLeuCysCysLys. 492

1643 ATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTG 1692
          :::|||   |||||||||   |||       ||||||:::||||||
  493 .....LysCysLysPheGlnProMetGlyThrValCysArgGluAlaVal 507

1693 AACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCC 1742
     |||:::|||||||||   |||   |||:::|||:::|||:::||||||
  508 AsnAspCysAspIleArgGluThrCysSerGlyAsnSerSerGlnCysAl 524

1743 ACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCC 1792
      ||||||:::||||||:::||||||||||:::|||:::      ||||||
  524 aProAsnIleHisLysMetAspGlyTyrSerCysAspGlyValGlnGlyI 541

1793 GCTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATC 1842
       |||:::::|||:::|||||||||||||||:::||||||:::||||||
  541 leCysPheGlyGlyArgCysLysThrArgAspArgGlnCysLysTyrIle 557

1843 TGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAA 1892
     ||||||   |||::::::::::|||||||||:::||||||||||||||||
  558 TrpGlyGlnLysValThrAlaSerAspLysTyrCysTyrGluLysLeuAs 574

1893 TACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGA 1942
     |   |||||||||||||||||||||||||||||||   |||   ||||
  574 nIleGluGlyThrGluLysGlyAsnCysGlyLysAspLysAspThrTrpI 591

1943 TTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAAT 1992
     ||||||||:::|||:::|||||||:::|||||||||||||||||||||
  591 leGlnCysAsnLysArgAspValLeuCysGlyTyrLeuLeuCysThrAsn 607

1993 CTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAAC 2042
     :::   :::   ||||||:::|||:::|||:::|||||||||      ||
  608 IleGlyAsnIleProArgLeuGlyGluLeuAspGlyGluIleThrSerTh 624

2043 TTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAG 2092
     |        :::|||||||||:::::::::|||||||||:::||||||
  624 rLeuValValGlnGlnGlyArgThrLeuAsnCysSerGlyGlyHisValL 641

2093 TTTTAGATGATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGT 2142
       |||:::::|||:::|||:::||||||||||||||||||||||||||||
  641 ysLeuGluGluAspValAspLeuGlyTyrValGluAspGlyThrProCys 657

2143 GGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCT 2192
     ||||||:::|||||||||||:::::::::|||||   :::   :::::
  658 GlyProGlnMetMetCysLeuGluHisArgCysLeuProValAlaSerPh 674

2193 AAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCC 2242
     :|||:::|||:::|||      :::::::::|||   :::|||||||||:
  674 eAsnPheSerThrCysLeuSerSerLysGluGlyThrIleCysSerGlyA 691

2243 ATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCA 2292
     ::|||||||||||||||||      |||:::|||:::      |||
  691 snGlyValCysSerAsnGluLeuLysCysValCysAsnArgHisTrpIle 707
```

```
  2293 GGGACAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAA 2342
       |||:::||||||                    |||   :::   |||
   708 GlySerAspCys.....................AsnThrTyrPheProHi 717

  2343 GGATGAAGGACCCAAG.:.....................GGTCCTAGTG 2368
         ::::::      |||                      |||   ::::
   717 sAsnAspAspAlaLysThrGlyIleThrLeuSerGlyAsnGlyValAlaG 734

  2369 CCACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCT 2418
       ::||||||:::||||||||   |||||||||:::||||||||    ||
   734 lyThrAsnIleIleIleGlyIleIleAlaGlyThrIleLeuValLeuAla 750

  2419 ATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAAATGTCAAGAAGAGAAG 2468
       ::::::||||||   |||:::||||||:::||||||   :::::::::||
   751 LeuIleLeuGlyIleThrAlaTrpGlyTyrLysAsnTyrArgGluGlnAr 767

   469 G 2469
       |
   767 g 767
```

```
seq_name: sp_human:O75078

seq_documentation_block:
ID   O75078      PRELIMINARY;      PRT;   769 AA.
AC   O75078;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MDC/ADAM11.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=98359734; PubMed=9693107;
RA   Sagane K., Ohya Y., Hasegawa Y., Tanaka I.;
RT   "Metalloproteinase-like, disintegrin-like, cysteine-rich proteins MDC2
RT   and MDC3: novel human cellular disintegrins highly expressed in the
RT   brain.";
RL   Biochem. J. 334:93-98(1998).
DR   EMBL; AB009675; BAA32352.1; -.
DR   HSSP; P18619; 1FVL.
DR   MEROPS; M12.976; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   ProDom; PD000664; -; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   SMART; SM00050; DISIN; 1.
SQ   SEQUENCE   769 AA;  83418 MW;  EEB091EB6730AC36 CRC64;


alignment_scores:
            Quality: 1983.00             Length:    817
              Ratio:  3.455               Gaps:     17
 Percent Similarity:  70.257   Percent Identity: 48.715

alignment_block:
 US-09-634-252A-2 x O75078   ..

 Align seg 1/1  to: O75078  from: 1  to: 769

    136 CTGCTTCTCGTCCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCC 185
        |||||||||   ||||||      |||   |||:::::::::::::
     12 LeuLeuLeuSerLeuLeuProThrProGlyLeuGlyThrGlnGlyProAl 28
```

```
  186 CCGCGCC......TGGGGGGCT............GCTGCGCCCAGCGCTC 217
          |||      ||||||:::            ::::::|||:::||||
   28 aGlyAlaLeuArgTrpGlyGlyLeuProGlnLeuGlyGlyProGlyAlaP 45

  218 CGCATTGGAATGAAACTGCAGAAAAAAATTTGGGAGTCCTGGCAGATGAA 267
      ||
   45 ro.............................................. 45

  268 GACAATACATTGCAACAGAATAGCAGCAGTAATATCAGTTACAGCAATGC 317

   45 ..............................................  45

  318 AATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACATCAACC 367
              |||:::|||   |||||||||||:::          :::::
   46 ..........GluValThrGluProSerArgLeuValArgGlu...SerS 58

  368 AAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAG 417
      ::   :::|||       :::   |||||||||:::::::|||::::::
   58 erGlyGlyGluValArgLysGlnGlnLeuAspThrArgValArgGlnGlu 74

  418 CAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGA 467
              :::      ||||||||||||||:::||||||   |||
   75 ProProGlyGlyProProValHisLeuAlaGlnValSerPheValIlePr 91

  468 AGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGT 517
       ||||||:::|||:::|||   |||||||||   ||||||:::   ||||
   91 oAlaPheAsnSerAsnPheThrLeuAspLeuGluLeuAsnHisHisLeuL 108

  518 TGTCTTCTGATTATGTGGAGATTCACTAC...GAAAATGGGAAACCACAG 564
      ||||||||:::|||||||||   |||:::   :::::::|||      |||
  108 euSerSerGlnTyrValGluArgHisPheSerArgGluGlyThrThrGln 124

  565 TACTCTAAGGGT...GGAGAGCACTGTTACTACCATGGAAGCATCAGAGG 611
      :::|||   |||   |||:::||||||||||||:::|||::::::|||||
  125 HisSerThrGlyAlaGlyAspHisCysTyrTyrGlnGlyLysLeuArgGl 141

  612 CGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCA 661
      |         |||   :::||||||||||||||:::||||||||||||:
  141 yAsnProHisSerPheAlaAlaLeuSerThrCysGlnGlyLeuHisGlyV 158

  662 TGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTT 711
      ::|||:::|||   ::::::::::|||:::::::||||||   |||::::::
  158 alPheSerAspGlyAsnLeuThrTyrIleValGluProGlnGluValAla 174

  712 ......CATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAAC 755
                        :::          ||||||:::|||   :::||
  175 GlyProTrpGlyAlaProGlnGlyProLeuProHisLeuIleTyrArgTh 191

  756 CTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAG 805
      |
  191 r.............................................. 191

  806 GTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAA..... 846
            |||:::|||   :::          :::
  192 ...........ProLeuLeuProAspProLeuGlyCysArgGluProGly 204

  846 ..............................................  846

  205 CysLeuPheAlaValProAlaGlnSerAlaProProAsnArgProArgLe 221

  847 .AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT.........ATATTTG 886
       ||||||||||||   |||:::      ||||||         :::   :
  221 uArgArgLysArgGlnValArg......ArgGlyHisProThrValHisS 236

  887 AAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAACGTAT 936
      ::|||   ||||||:::||||||:::::::::||||||||:::   :::
  236 erGluThrLysTyrValGluLeuIleValIleAsnAspHisGlnLeuPhe 252

  937 AAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGT 986
      ::::::   |||:::|||   :::   |||:::||||||||||||||||
  253 GluGlnMetArgGlnSerValValLeuThrSerAsnPheAlaLysSerVa 269
```

```
 987 GGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTG 1036
     ||||||||||:::|||   ||||||||||||||||||||||||||:::|
 269 lValAsnLeuAlaAspValIleTyrLysGluGlnLeuAsnThrArgIleV 286

1037 TCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCACC 1086
     ||||||||||:::|||||||||:::::   |||:::|||::::::
 286 alLeuValAlaMetGluThrTrpAlaAspGlyAspLysIleGlnValGln 302

1087 ACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGC.. 1134
        :::   ::::::   |||   ::::::      ||||||::::::
 303 AspAspLeuLeuGluThrLeuAlaArgLeuMetValTyrArgArgGluGl 319

1135 .ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTC 1183
      :::   :::   :::|||||:::|||||:::|||      ||||||:
 319 yLeuProGluProSerAspAlaThrHisLeuPheSerGlyArgThrPheG 336

1184 ACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACA 1233
     ::          |||:::   :::|||   ||||||:::||||||   :::
 336 lnSerThrSerSerGlyAlaAlaTyrValGlyGlyIleCysSerLeuSer 352

 234 AGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACAAGT 1283
     :::|||   ||||||||||||||||||       |||:::|||   ::
 353 HisGlyGlyGlyValAsnGluTyrGlyAsnMetGlyAlaMetAlaValTh 369

1284 ATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGG...GAACCTT 1330
     :|||:::|||:::|||:::||||||||||:::::|||   :::
 369 rLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMetMetTrpAsnLysHisA 386

1331 CTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATC 1380
      |||              |||   |||   :::   |||   |||||||||
 386 rgSerSerAlaGlyAspCysLysCysProAspIleTrpLeuGlyCysIle 402

1381 ATGGAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCAT 1430
     ||||||:::||||||           ||||||||||||:::||||||||
 403 MetGluAspThrGlyPheTyrLeuProArgLysPheSerArgCysSerIl 419

1431 TTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCA 1480
     |   ||||||:::::::|||||||:::||||||||:::||||||||||
 419 eAspGluTyrAsnGlnPheLeuGlnGluGlyGlyGlySerCysLeuPheA 436

1481 ACAGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATACGTG 1530
     ||:::|||   ||||||:::::::|||   ||||||||||||||:::|||
 436 snLysProLeuLysLeuLeuAspProProGluCysGlyAsnGlyPheVal 452

1531 GAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTAT..... 1575
     ||||||||||||||||||||||||||||         ||||||
 453 GluAlaGlyGluGluCysAspCysGlySerValGlnGluCysSerArgAl 469

1576 ....GGATTATGCTGTAAGAAATGTTCCCTCTCCAACGGGGCTCACTGCA 1621
         |||   ||||||||||||||:::|||::::::   |||   ||||
 469 aGlyGlyAsnCysCysLysLysCysThrLeuThrHisAspAlaMetCysS 486

1622 GCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGG 1671
     ||||||||   ||||||::::::      |||   ::::::||||||||
 486 erAspGlyLeuCysCysArgArg......CysLysTyrGluProArgGly 500

1672 TATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTAC 1721
        :::||||||:::||||||||||||||||||||:::|||   |||||
 501 ValSerCysArgGluAlaValAsnGluCysAspIleAlaGluThrCysTh 517

1722 TGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATG 1771
     ||||||||||:::||||||||||||||||||||||   |||||||||
 517 rGlyAspSerSerGlnCysProProAsnLeuHisLysLeuAspGlyTyrT 534

1772 CATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGA 1821
       |||::::::::::|||||||||||||||:::|||:::|||||||||||
 534 yrCysAspHisGluGlnGlyArgCysTyrGlyGlyArgCysLysThrArg 550

1822 GACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAA 1871
     |||:::|||||||||   :::||||||      ||||||      |||::
 551 AspArgGlnCysGlnValLeuTrpGlyHisAlaAlaAla......AspAr 565

1872 GTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCG 1921
```

```
          :|||||||||||||||||||:::|||||||||||:::|||:::||||
  565 gPheCysTyrGluLysLeuAsnValGluGlyThrGluArgGlySerCysG 582

 1922 GGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGT 1971
      ||:::   |||:::   |||:::||||||||||||:::|||||||:::|||
  582 lyArgLysGlySerGlyTrpValGlnCysSerLysGlnAspValLeuCys 598

 1972 GGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACT 2021
      ||||||||||||||:::|||::::::   ||||||||:::|||:::||
  599 GlyPheLeuLeuCysValAsnIleSerGlyAlaProArgLeuGlyAspLe 615

 2022 TCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTG 2071
      |   |||:::|||      ::::::||||||||||||||||:::   :::|
  615 uValGlyAspIleSerSerValThrPheTyrHisGlnGlyLysGluLeuA 632

 2072 ACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTAT 2121
      |||||   |||:::||||||   |||   |||   :::|||::::::|||
  632 spCysArgGlyGlyHisValGlnLeuAlaAspGlySerAspLeuSerTyr 648

 2122 GTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAA 2171
      ||||||||||||||   |||||||||:::|||:::|||||||||:::::
  649 ValGluAspGlyThrAlaCysGlyProAsnMetLeuCysLeuAspHisAr 665

 2172 GTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCCA 2221
      :||||||      :::|||:::|||:::|||:::||||||   :::::::
  665 gCysLeuProAlaSerAlaPheAsnPheSerThrCysProGlySerGlyG 682

 2222 AGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGC 2271
      ::   ::::::||||||   ||||||||||||||||||:::   |||
  682 luArgArgIleCysSerHisHisGlyValCysSerAsnGluGlyLysCys 698

 2272 ATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGT 2321
      ||||||:::      |||:::|||   ||||||||||||:::::|||::
  699 IleCysGlnProAspTrpThrGlyLysAspCysSerIleHisAsnProLe 715

 2322 TAGGAACCTTCACCCCCCC...AAGGATGAAGGACCCAAGGGTCCTAGTG 2368
      :   :::      |||      :::   |||      ||||||||||||:
  715 uProThrSerProProThrGlyGluThrGluArgTyrLysGlyProSerG 732

 2369 CCACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCT 2418
      ::|||||:::||||||||||||||||||||||||:::||||||||||||
  732 lyThrAsnIleIleIleGlySerIleAlaGlyAlaValLeuValAlaAla 748

 2419 ATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAAATGTCAAGAAGAGAAG 2468
      ||||||||||||||||||||||||||||||||||||:::::::::  ||
  749 IleValLeuGlyGlyThrGlyTrpGlyPheLysAsnIleArgArgGlyAr 765

 2469 G 2469
      |
  765 g 765
```

```
seq_name: sp_rodent:Q9R1V5

seq_documentation_block:
ID   Q9R1V5      PRELIMINARY;      PRT;   821 AA.
AC   Q9R1V5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   ADAM22 BETA.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=99365303; PubMed=10433968;
RA   Sagane K., Yamazaki K., Mizui Y., Tanaka I.;
RT   "Cloning and chromosomal mapping of mouse ADAM11,ADAM22 and ADAM23.";
RL   Gene 236:79-86(1999).
DR   EMBL; AB009674; BAA83383.1; -.
DR   HSSP; P18619; 1FVL.
```

```
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PRINTS; PR00289; DISINTEGRIN.
DR   ProDom; PD000664; -; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   SMART; SM00050; DISIN; 1.
SQ   SEQUENCE   821 AA;  90952 MW;  4F8F12B425EB683C CRC64;


    ment_scores:
           Quality: 1971.00              Length:    739
             Ratio:  3.539                Gaps:     11
 Percent Similarity:  75.372   Percent Identity: 50.474

alignment_block:
 US-09-634-252A-2 x Q9R1V5   ..

 Align seg 1/1  to: Q9R1V5  from: 1  to: 821

     319 ATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACATCAACCA 368
          ::::::::::::      :::|||   ||||||||||||   ::::::
      43 LeuGluArgGlnSerIleIleProLeuArgLeuIleTyrArgLeuGlyGl 59

     369 AGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGC 418
          ::::::|||:::   ::::::   |||||||||:::::|||   :::
      59 yGluAspGluThrGlnHisAsnGlnLeuAspThrArgValArgGlyAspP 76

     419 AAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAA 468
                    :::   :::|||:::   :::|||||||||:::::::::
      76 roGlyGlyProGlnLeuThrHisValAspLysAlaSerPheArgValAsp 92

     469 GCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTCTT 518
          |||||||||:::::::|||:::|||||:::::::|||||:::   |||||
      93 AlaPheGlyThrSerPheValLeuAspValLeuLeuAsnHisGluLeuLe 109

     519 GTCTTCTGATTATGTGGAGATTCACTACGAAAAT...GGGAAACCACAGT 565
          |||||||   |||||||||   :::   |||:::   ||||||
     109 uSerSerGlyTyrValGluArgGlnIleGluHisGlyGlyLysValValG 126

     566 ACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGCGTC 515
          :::|||||||||||||||||||||||:::|||:::|||||||||
     126 luAsnLysGlyGlyGluHisCysTyrTyrGlnGlyGlnIleArgGlyAsn 142

     616 AAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTT 665
               |||   |||||||||||||||||:::|||||||||||||||||
     143 ProValSerPheValAlaLeuSerThrCysHisGlyLeuHisGlyMetPh 159

     666 TGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATG 715
          |   |||   :::   :::|||:::|||||||||   |||         :
     159 eTyrAspGlyAsnHisThrTyrLeuIleGluProGluGlu.........A 173

     716 ATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCAGGA 765
          ::|||||||||             |||      :::          :::
     173 snGluLysSerGlnGluSerSerHisCysHisSerValTyrLysSerArg 189

     766 CAGTATTCTAAGCAAATGAAGAATCTCACTATGGAA......AGAGGTGA 809
          |||::::::      :::   :::|||      |||      |||   ::
     190 GlnPheGluPheProLeuAspAspLeuProSerGluPheGlnArgValAs 206

     810 CCAGTGGCCC......TTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGA 853
          :      |||      |||:::   :::   :::      :::|||::::
     206 nIleThrProProGlnPheIleLeuLysProArgLeuLysArgArgLysA 223

     854 AGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATTTG 903
          :::::   ::::::      |||::::::   ||||||   ||||||:::
     223 rgGlnLeuLeuArgPheProArgAsnValGluGluGluThrLysTyrIle 239

     904 GAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTC 953
          ||||||||||||||||||||||||      :::|||||||||||||   ||
     240 GluLeuMetIleValAsnAspHisLeuMetPheLysLysHisArgLeuSe 256

     954 TCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATT 1003
          |   ::::::||||||::::::|||||||||||||||||||:::::|||
     256 rValValTyrThrAsnThrTyrAlaLysSerValValAsnMetAlaAspV 273

    1004 CTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAG 1053
          |||||||||:::||||||:::||||||||:::|||||||||||||:::|||
     273 alIleTyrLysAspGlnLeuLysThrArgIleValLeuValAlaMetGlu 289

    1054 ACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAGAT 1103
          |||||||:::      :::::::::   |||:::   |||||||:::
     290 ThrTrpAlaAlaAspAsnLysPheAlaIleSerGluAsnProLeuIleTh 306

    1104 GCTCCATGAGTTCTCAAAATACCGGCAGCGC...ATTAAGCAGCATGCTG 1150
          |||:::||||||   ||||||||:::      ||||||:::   ||||
     306 rLeuArgGluPheMetLysTyrArgArgAspPheIleLysGluLysAlaA 323

    1151 ATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGT 1200
          ||||||||||||||:::|||         |||:::   :::||||||:::
     323 spAlaValHisLeuPheSerGlySerGlnPheGluSerSerArgSerGly 339

    1201 CTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAA 1250
          :::|||:::||||||:::||||||      ||||||   ||||||||
     340 AlaAlaTyrIleGlyGlyIleCysSerLeuLeuArgGlyGlyGlyValAs 356

    1251 TGAGTATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTGG 1300
          ||||:::|||              :::|||   :::|||:::||||||||||
     356 nGluPheGlyLysThrAspLeuMetAlaValThrLeuAlaGlnSerLeuA 373

    1301 CTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAA... 1347
          ||:::|||:::||||||               |||   |||   |||
     373 laHisAsnValGlyIle.............IleSerAspLysArgLysLeu 385

    1348 .............TGTGACTGCACAGAATCCTGGGGTGGCTGCATCATGGA 1385
                      |||   |||   ::::::|||:::||||||||||||
     386 AlaSerGlyGluCysLysCysGluAspThrTrpSerGlyCysIleMetGl 402

    1386 GGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGG 1435
          :::||||||              :::||||||::::::|||::::::   |
     402 yAspThrGlyTyrTyrLeuProLysLysPheThrGlnCysAsnValGluG 419

    1436 AGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGG 1485
          |||||:::||||||||::   ||||||||||||||||||||||||:::
     419 luTyrHisAspPheLeuAsnSerGlyGlyGlyAlaCysLeuPheAsnLys 435

    1486 CCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATACGTGGAAGC 1535
          |||:::||||||::::::|||   |||||||||||||||:::::|||::
     436 ProSerLysLeuLeuAspProProGluCysGlyAsnGlyPheIleGluTh 452

    1536 TGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTAT.........G 1576
          :|||||||||||||||||||      :::||||||             :
     452 rGlyGluGluCysAspCysGlyThrProAlaGluCysAlaLeuGluGlyA 469

    1577 GATTATGCTGTAAGAAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGAC 1626
          ::   ||||||||||||||:::|||::::::   ::::::|||||||||
     469 laGluCysCysLysLysCysThrLeuThrGlnAspSerGlnCysSerAsp 485

    1627 GGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGA 1676
          |||   ||||||:::       :::|||   ||||||||   |||
     486 GlyLeuCysCysLys......LysCysLysPheGlnProLeuGlyThrVa 500

    1677 ATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTACTGGAG 1726
          ||||||:::|||||||||:::|||||||||   |||   |||:::|||:
     500 lCysArgGluAlaValAsnAspCysAspIleArgGluIleCysSerGlyA 517

    1727 ACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATGCATGC 1776
          ::|||:::||||||   ||||||:::|||||||:::||||||||||:::|||
     517 snSerSerGlnCysAlaProAsnValHisLysMetAspGlyTyrSerCys 533
```

```
1777 AATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAA 1826
      :::   :::||||||   |||:::::::|||:::|||||||||||||||::
 534 AspGlyThrGlnGlyIleCysPheGlyGlyArgCysLysThrArgAspAr 550

1827 CCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCT 1876
      :||||||:::|||||||||||   |||:::::::::||||||:::::|
 550 gGlnCysLysTyrIleTrpGlyGlnLysValThrAlaSerAspArgTyrC 567

1877 GCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAG 1926
      ||||||||||||||||   |||||||||||||||||||||||||||||
 567 ysTyrGluLysLeuAsnIleGluGlyThrGluLysGlyAsnCysGlyLys 583

1927 GATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGTGGATT 1976
      |||   |||   |||   ||||||:::|||:::||||||:::||||||::
 584 AspLysAspThrTrpThrGlnCysAsnLysArgAspValLeuCysGlyTy 600

1977 CTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGG 2026
      :|||||||||||||||:::   :::   ||||||:::|||:::|||:::|
 600 rLeuLeuCysThrAsnIleGlyAsnIleProArgLeuGlyGluLeuAspG 617

 027 GTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGC 2076
      ||||||||      |||         :::||||||||||:::::::::|||
 617 lyGluIleThrSerThrLeuValValGlnGlnGlyArgThrLeuAsnCys 633

2077 AGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTATGTAGA 2126
      ||||||||||||||   |||:::::::|||:::|||:::||||||||||
 634 SerGlyAlaHisValLysLeuGluGluAspValAspLeuGlyTyrValGl 650

2127 AGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCC 2176
      ||||||||||||||||||||||:::||||||||||||:::::::::||||
 650 uAspGlyThrProCysGlyProGlnMetMetCysLeuGluHisArgCysL 667

2177 TACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCCAAGGGT 2226
      ||   :::   ::::::|||:::|||:::|||      ::::::   |||
 667 euProValAlaSerPheAsnPheSerThrCysSerSerSerLysAlaGly 683

2227 AAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTG 2276
      ||||||||||||:::||||||||||||||||||      |||:::||
 684 ThrValCysSerGlyAsnGlyValCysSerAsnGluLeuLysCysValCy 700

2277 TGATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGTTAGGA 2326
      |:::      |||:::|||:::||||||                    :
 700 sAsnArgHisTrpThrGlyAlaAspCys.....................G 710

2327 ACCTTCACCCCCCCAAGGATGAAGGACCCAAG..................... 2358
      ::   |||   |||   ::::::     |||
 710 lyThrHisProHisAsnAspAspAlaLysThrGlyIleThrLeuSer 726

2359 ......GGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGGTGC 2402
            |||   ::::::||||||:::|||||||||   ||||||||::
 727 GlyAsnGlyValAlaGlyThrAsnIleIleIleGlyIleIleAlaGlyTh 743

2403 CATCCTGGTAGCAGCTATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAA 2452
      :|||||||||   |||::::::||||||   |||:::||||||:::||||
 743 rIleLeuValLeuAlaLeuIleLeuGlyIleThrAlaTrpGlyTyrLysA 760

2453 ATGTCAAGAAGAGAAGG 2469
      ||   ::::::::::|||
 760 snTyrArgGluGlnArg 765
```

seq_name: sp_rodent:Q9R1V6

seq_documentation_block:

```
ID   Q9R1V6      PRELIMINARY;      PRT;   857 AA.
AC   Q9R1V6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   ADAM22 ALPHA.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN;
RX   MEDLINE=99365303; PubMed=10433968;
RA   Sagane K., Yamazaki K., Mizui Y., Tanaka I.;
RT   "Cloning and chromosomal mapping of mouse ADAM11,ADAM22 and ADAM23.";
RL   Gene 236:79-86(1999).
DR   EMBL; AB009674; BAA83382.1; -.
DR   HSSP; P18619; 1FVL.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PRINTS; PR00289; DISINTEGRIN.
DR   ProDom; PD000664; -; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   SMART; SM00050; DISIN; 1.
SQ   SEQUENCE   857 AA;  94740 MW;  CB88FB7000208E09 CRC64;
```

alignment_scores:

Quality: 1971.00 Length: 739
Ratio: 3.539 Gaps: 11
Percent Similarity: 75.372 Percent Identity: 50.474

alignment_block:

US-09-634-252A-2 x Q9R1V6 ..

Align seg 1/1 to: Q9R1V6 from: 1 to: 857

```
 319 ATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACATCAACCA 368
      ::::::::::::      :::|||   ||||||||||||   ::::::
  43 LeuGluArgGlnSerIleIleProLeuArgLeuIleTyrArgLeuGlyGl 59

 369 AGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGC 418
      ::::::|||:::   ::::::   |||||||||:::::|||   :::
  59 yGluAspGluThrGlnHisAsnGlnLeuAspThrArgValArgGlyAspP 76

 419 AAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAA 468
                :::   :::|||:::   :::||||||||:::::::::
  76 roGlyGlyProGlnLeuThrHisValAspLysAlaSerPheArgValAsp 92

 469 GCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGTT 518
      |||||||||:::::|||:::||||||:::::||||||:::   |||||
  93 AlaPheGlyThrSerPheValLeuAspValLeuLeuAsnHisGluLeuLe 109

 519 GTCTTCTGATTATGTGGAGATTCACTACGAAAAT...GGGAAACCACAGT 565
      |||||||   |||||||||   :::   |||:::   ||||||
 109 uSerSerGlyTyrValGluArgGlnIleGluHisGlyGlyLysValValG 126

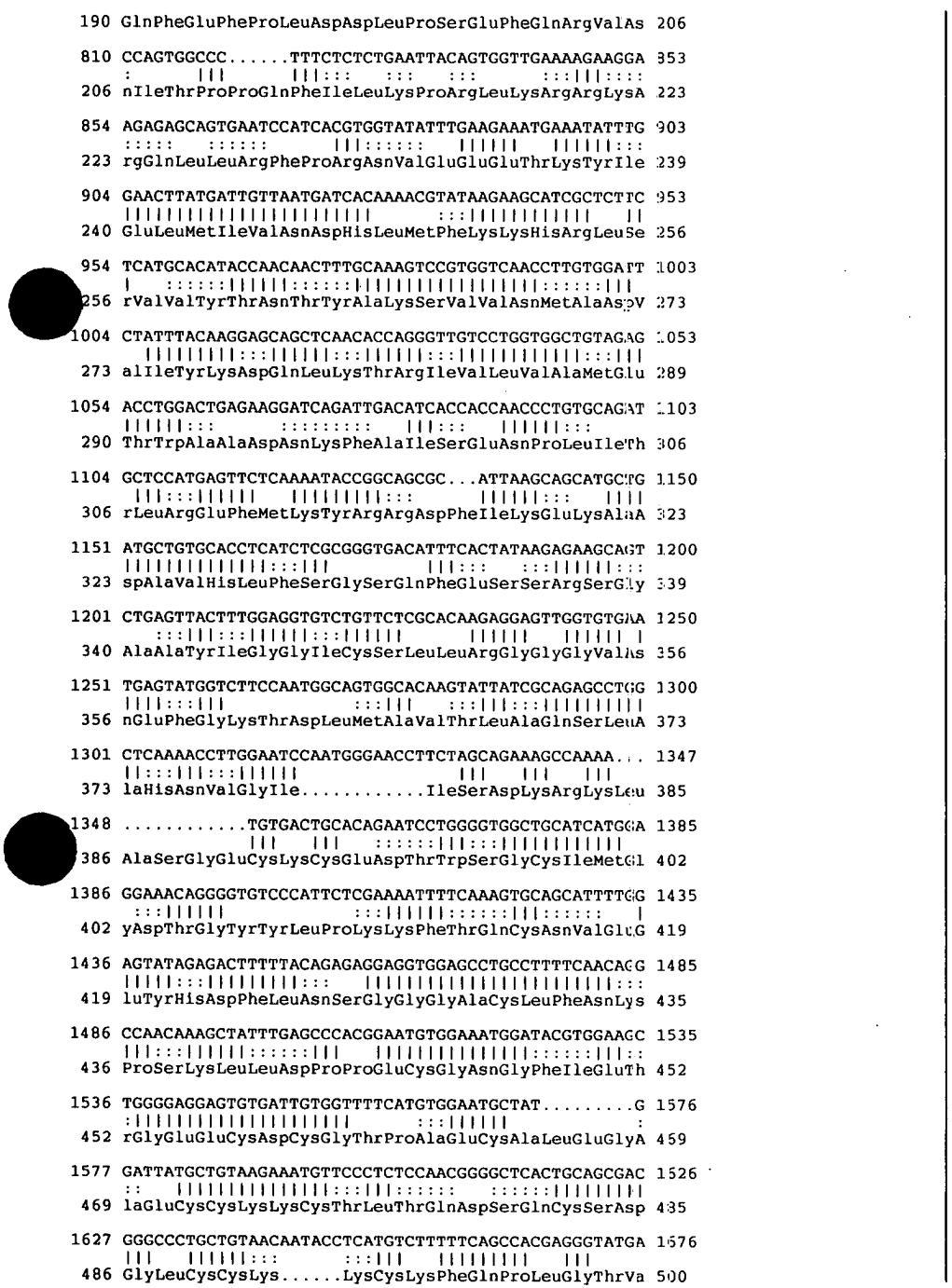
 566 ACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGCGTC 615
       :::|||||||||||||||||||||||||:::|||:::|||||||||
 126 luAsnLysGlyGlyGluHisCysTyrTyrGlnGlyGlnIleArgGlyAsn 142

 616 AAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTT 665
           |||   ||||||||||||||||:::|||||||||||||||||
 143 ProValSerPheValAlaLeuSerThrCysHisGlyLeuHisGlyMetPh 159

 666 TGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATG 715
      |   |||   :::   :::|||:::|||||||||   |||         :
 159 eTyrAspGlyAsnHisThrTyrLeuIleGluProGluGlu.........A 173

 716 ATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCAGGA 765
      ::|||||||||            |||      :::         :.::
 173 snGluLysSerGlnGluSerSerHisCysHisSerValTyrLysSerArg 189

 766 CAGTATTCTAAGCAAATGAAGAATCTCACTATGGAA......AGAGGTGA 809
      |||::::::      :::   :::|||      |||      |||   ::
```



```
 190 GlnPheGluPheProLeuAspAspLeuProSerGluPheGlnArgValAs 206

 810 CCAGTGGCCC......TTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGA 853
     :      |||      |||:::   :::   :::      :::|||:::
 206 nIleThrProProGlnPheIleLeuLysProArgLeuLysArgArgLysA 223

 854 AGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATTTG 903
     :::::   ::::::      |||:::::   ||||||   |||||||::
 223 rgGlnLeuLeuArgPheProArgAsnValGluGluGluThrLysTyrIle 239

 904 GAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTC 953
     ||||||||||||||||||||||||      :::||||||||||||   ||
 240 GluLeuMetIleValAsnAspHisLeuMetPheLysLysHisArgLeuSe 256

 954 TCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATT 1003
     |   :::::|||||||::::::|||||||||||||||||||:::::|||
 256 rValValTyrThrAsnThrTyrAlaLysSerValValAsnMetAlaAspV 273

1004 CTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAG 1053
       |||||||||:::||||||:::|||||||:::|||||||||||||:::|||
 273 alIleTyrLysAspGlnLeuLysThrArgIleValLeuValAlaMetGlu 289

1054 ACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAGAT 1103
     ||||||:::      :::::::::   |||:::   |||||||:::
 290 ThrTrpAlaAlaAspAsnLysPheAlaIleSerGluAsnProLeuIleTh 306

1104 GCTCCATGAGTTCTCAAAATACCGGCAGCGC...ATTAAGCAGCATGCTG 1150
       |||:::||||||   |||||||||:::      ||||||:::   ||||
 306 rLeuArgGluPheMetLysTyrArgArgAspPheIleLysGluLysAlaA 323

1151 ATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGT 1200
     ||||||||||||||:::|||         |||:::   :::|||||||:::
 323 spAlaValHisLeuPheSerGlySerGlnPheGluSerSerArgSerGly 339

1201 CTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAA 1250
       :::|||:::||||||:::||||||      ||||||   |||||| |
 340 AlaAlaTyrIleGlyGlyIleCysSerLeuLeuArgGlyGlyGlyValAs 356

1251 TGAGTATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTGG 1300
     ||||:::|||            :::|||   :::|||:::||||||||||
 356 nGluPheGlyLysThrAspLeuMetAlaValThrLeuAlaGlnSerLeuA 373

1301 CTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAA... 1347
     ||:::|||:::||||||            |||   |||   |||
 373 laHisAsnValGlyIle.............IleSerAspLysArgLysLeu 385

1348 ............TGTGACTGCACAGAATCCTGGGGTGGCTGCATCATGGA 1385
                |||   |||   ::::::|||:::||||||||||||
 386 AlaSerGlyGluCysLysCysGluAspThrTrpSerGlyCysIleMetGl 402

1386 GGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGG 1435
      :::||||||            :::||||||:::::::|||::::::   |
 402 yAspThrGlyTyrTyrLeuProLysLysPheThrGlnCysAsnValGluG 419

1436 AGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACACG 1485
     |||||:::||||||||:::   ||||||||||||||||||||||||:::
 419 luTyrHisAspPheLeuAsnSerGlyGlyGlyAlaCysLeuPheAsnLys 435

1486 CCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATACGTGGAAGC 1535
     |||:::||||||::::::|||   ||||||||||||||:::::|||::
 436 ProSerLysLeuLeuAspProProGluCysGlyAsnGlyPheIleGluTh 452

1536 TGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTAT.........G 1576
     :||||||||||||||||||      :::||||||            :
 452 rGlyGluGluCysAspCysGlyThrProAlaGluCysAlaLeuGluGlyA 469

1577 GATTATGCTGTAAGAAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGAC 1526
     ::   ||||||||||||||:::|||::::::   ::::::|||||||||
 469 laGluCysCysLysLysCysThrLeuThrGlnAspSerGlnCysSerAsp 485

1627 GGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGA 1676
     |||   ||||||:::      :::|||   |||||||||   |||
 486 GlyLeuCysCysLys......LysCysLysPheGlnProLeuGlyThrVa 500

1677 ATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTACTGGAG 1726
      ||||||:::|||||||||:::|||||||||   |||   |||:::|||:
 500 lCysArgGluAlaValAsnAspCysAspIleArgGluIleCysSerGlyA 517

1727 ACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATGCATGC 1776
     ::|||:::||||||   ||||||:::||||||:::|||||||||:::|||
 517 snSerSerGlnCysAlaProAsnValHisLysMetAspGlyTyrSerCys 533

1777 AATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAA 1826
     :::   :::||||||   |||:::::|||:::|||||||||||||||:.
 534 AspGlyThrGlnGlyIleCysPheGlyGlyArgCysLysThrArgAspAr 550

1827 CCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCT 1876
     :||||||:::|||||||||||||   |||:::::::::||||||::::::|
 550 gGlnCysLysTyrIleTrpGlyGlnLysValThrAlaSerAspArgTyrC 567

1877 GCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAG 1926
     ||||||||||||||||   |||||||||||||||||||||||||||||||
 567 ysTyrGluLysLeuAsnIleGluGlyThrGluLysGlyAsnCysGlyLys 583

1927 GATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGTGGATT 1976
     |||   |||   |||   ||||||:::|||:::||||||:::||||||::
 584 AspLysAspThrTrpThrGlnCysAsnLysArgAspValLeuCysGlyTy 600

1977 CTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGG 2026
     :||||||||||||||:::   :::   ||||||:::|||:::|||:::|
 600 rLeuLeuCysThrAsnIleGlyAsnIleProArgLeuGlyGluLeuAspG 617

2027 GTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGC 2076
     ||||||||      |||         :::|||||||||:::::::::|||
 617 lyGluIleThrSerThrLeuValValGlnGlnGlyArgThrLeuAsnCys 633

2077 AGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGGCTATGTAGA 2126
     ||||||||||||||   |||:::::|||:::|||:::||||||||||||
 634 SerGlyAlaHisValLysLeuGluGluAspValAspLeuGlyTyrValGl 650

2127 AGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCC 2176
     ||||||||||||||||||||:::|||||||||||:::::::::||||
 650 uAspGlyThrProCysGlyProGlnMetMetCysLeuGluHisArgCysL 667

2177 TACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCCAAGGGT 2226
     ||   :::   ::::::|||:::|||:::|||      ::::::   |||
 667 euProValAlaSerPheAsnPheSerThrCysSerSerSerLysAlaGly 683

2227 AAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTG 2276
        ||||||||||||:::||||||||||||||||      |||:::||
 684 ThrValCysSerGlyAsnGlyValCysSerAsnGluLeuLysCysValCy 700

2277 TGATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGTTAGGA 2326
     |:::      |||:::|||:::||||||                    :
 700 sAsnArgHisTrpThrGlyAlaAspCys.....................G 710

2327 ACCTTCACCCCCCCAAGGATGAAGGACCCAAG.................. 2358
     ::   |||   |||   ::::::      |||
 710 lyThrHisProHisAsnAspAspAlaLysThrGlyIleThrLeuSer 726

2359 ......GGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGGTGC 2402
           |||   ::::::||||||::||||||||   ||||||||::
 727 GlyAsnGlyValAlaGlyThrAsnIleIleIleGlyIleIleAlaGlyTh 743

2403 CATCCTGGTAGCAGCTATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAA 2452
     :|||||||||   |||:::::|||||   |||:::|||||:::||||
 743 rIleLeuValLeuAlaLeuIleLeuGlyIleThrAlaTrpGlyTyrLysA 760

2453 ATGTCAAGAAGAGAAGG 2469
     ||   :::::::::|||
 760 snTyrArgGluGlnArg 765
```

seq_name: sp_vertebrate:O42596

seq_documentation_block:
ID O42596 PRELIMINARY; PRT; 935 AA.

```
AC   O42596;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   METALLOPROTEASE-DISINTEGRIN MDC11B (METALLOPROTEASE/DISINTEGRIN
DE   XMDC11.2).
GN   MDC11B.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cai H., Kraetzschmar J., Alfandari D., Hunnicutt G., Blobel C.P.;
RT   "Neural crest-specific and general expression of distinct
RT   metalloprotease-disintegrins in early Xenopus laevis development.";
RL   Dev. Biol. 0:0-0(1998).
RN   [2]
RP   SEQUENCE OF 464-511 FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE=97349132; PubMed=9205136;
RA   Shilling F.M., Kratzschmar J., Cai H., Weskamp G., Gayko U.,
RA   Leibow J., Myles D.G., Nuccitelli R., Blobel C.P.;
RT   "Identification of metalloprotease/disintegrins in Xenopus laevis
RT   testis with a potential role in fertilization.";
RL   Dev. Biol. 186:155-164(1997).
CC   -!- SIMILARITY: CONTAINS A DISINTEGRIN DOMAIN.
DR   EMBL; AF032383; AAC61847.1; -.
DR   EMBL; U78188; AAB87148.1; -.
DR   HSSP; P18619; 1FVL.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PRINTS; PR00289; DISINTEGRIN.
DR   ProDom; PD000664; -; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   SMART; SM00050; DISIN; 1.
KW   Hydrolase; Metalloprotease; Protease.
SQ   SEQUENCE   935 AA;  104161 MW;  77B7AFDCC5C77C90 CRC64;


alignment_scores:
          Quality: 1944.00          Length:    793
            Ratio:  3.411            Gaps:     14
nt Similarity:  71.879   Percent Identity: 48.676

alignment_block:
 US-09-634-252A-2 x O42596   ..

 Align seg 1/1  to: O42596  from: 1  to: 935

     172 GCCTCGTCCCGGCCCCGCGCCTGGGGGGCTGCTGCGCCCAGCGCTCCGCA 221
         ||||||:::||||||:::      |||:::::::::   :::
      24 AlaSerThrArgProGluAsnGlyGlyThrSerGlyMetGln........ 37

     222 TTGGAATGAAACTGCAGAAAAAAATTTGGGAGTCCTGGCAGATGAAGACA 271
                       :::|||:::   :::||||||:::   ||||||
      38 .............ArgLysLysGluAsnSerValLeuGlyMetGluAsp. 49

     272 ATACATTGCAACAGAATAGCAGCAGTAATATCAGTTACAGCAATGCAATG 321

      49 ..................................................... 49

     322 CAGAAAGAAATCACACTGCCTTCAAGACTCATATATTACATCAACCAAGA 371
                     |||:::|||   |||||||||:::      ::::::||

      50 ............ThrValProLeuArgLeuIlePheSerAsnGluGluAs 62

     372 CTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGA......CACC 415
         |:::::::::          :::|||:::|||:::::::|||
      62 pAsnGlnThrThrGlnGlyLeuLeuSerThrArgValArgAlaGlySerP 79

     416 AGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATT 465
           |||   :::::::::   :::|||:::||||||||||||||||||||
      79 roGlnHisGlnAspGlnLeuThrHisValAlaGlnAlaSerPheGlnIle 95

     466 GAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATACTGAACAATGGTTT 515
         :::||||||||||||:::||||||||||||:::   ||||||:::   ||
      96 AspAlaPheGlySerSerPheIleLeuAspValGluLeuAsnHisAspLe 112

     516 GTTGTCTTCTGATTATGTGGAGATTCACTAC...GAAAATGGGAAACCAC 562
         ||||||||||||||||   |||   |||      ::::::||||||
     112 uLeuSerSerAspTyrArgGluArgHisValThrGlnAspGlyLysThrV 129

     563 AGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGC 612
                 |||||||||||||||||||||||:::|||:::|||||||||
     129 alGluValLysGlyGlyGluHisCysTyrTyrGlnGlyGlnIleArgGly 145

     613 GTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCAT 662
               |||   ||||||||||||||||||||||||||||||||||||
     146 LysAlaLysSerPheValAlaLeuSerThrCysAsnGlyLeuHisGlyMe 162

     663 GTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTC 712
         ||||   |||   :::   :::|||:::|||||||||   |||      :
     162 tPheCysAspGlyAsnHisThrTyrLeuIleGluProGlyGluLysTyrA 179

     713 ATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCA 762
         ::   :::::::::
     179 snProAsnGluAspTyr..................................... 184

     763 GGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAA.......... 801
           |||:::   :::      |||:::   ::::::|||
     185 ...GlnPheHisSerValTyrLysSerLysValLeuGluPheProLeuAs 200

     802 .......AGAGGTGACCAGTGGCCCTTT.................CTCT 826
                 ::::::   |||   :::                  ||||
     200 pGluLeuProSerGluPheTrpAlaLeuAsnAspThrSerValArgLeuS 217

     827 CTGAATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCA...TCA 873
         ||:::            :::|||:::||||||   ::::::
     217 erGlnGlnThrArgSerGlnArgLysLysArgGlnThrArgArgTyrPro 233

     874 CGTGGTATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAATGA 923
         |||::::::   :::|||   ||||||:::|||||||||||||||||||
     234 ArgAsnValGluAspGluThrLysTyrValGluLeuMetIleValAsnAs 250

     924 TCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAACAACT 973
         ||||      ||||||||||||||||   |||   :::|||||||||::::
     250 pHisLeuMetTyrLysLysHisArgLeuSerValGlyHisThrAsnSerT 267

     974 TTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTC 1023
         ::|||||||||||||||||:::::|||   ||||||||||||||||||
     267 yrAlaLysSerValValAsnMetAlaAspLeuIleTyrLysGluGlnLeu 283

    1024 AACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCA 1073
         |||||||||:::||||||||||||:::||||||||:::      :::::
     284 AsnThrArgIleValLeuValAlaMetGluThrTrpAlaThrAspAsnLy 300

    1074 GATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAAT 1123
         :::::::|||:::   ||||||:::      |||   ||||||   ||||
     300 sPheSerIleSerGluAsnProLeuLeuThrLeuLysGluPheMetLysT 317

    1124 ACCGGCAGCGC...ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCG 1170
         |||||:::      ||||||:::   :::||||||||||||||:::|||
     317 yrArgArgAspPheIleLysAspLysSerAspAlaValHisLeuPheSer 333

    1171 CGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGT 1220
               ::::::                |||      :::
     334 GlySerGlnLeuArgValAlaAlaValValLeuArgIleLeuValGluTr 350
```

```
1221 CTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGG 1270
     ||||||      :::|||   |||||||||||:::|||   |||    :
 350 pCysSerLeuLeuLysGlyGlyGlyValAsnGluPheGlyLysProAspV 367

1271 CAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAA 1320
     :::::|||   :::|||:::||||||||||||:::||||||||||||
 367 alMetAlaValThrLeuAlaGlnSerLeuAlaHisAsnLeuGlyIlePhe 383

1321 TGGGAACCTTCTAGCAGAAAGCCA.............AAATGTGACTGCAC 1358
            |||:::::::|||                 :::|||   |||
 384 .........SerAspLysLysLysLeuLeuSerGlyGluCysLysCysGl 397

1359 AGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCATTCTC 1408
     ::::::|||:::||||||||||||   :::   |||
 397 uAspThrTrpSerGlyCysIleMetGlyAspIleGlyTyrTyrLeuProS 414

1409 GAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGA 1458
       |||||||||   |||:::|||   ||||||:::::|||||::::::
 414 erLysPheSerValCysAsnIleGluGluTyrHisGluPheLeuAsnAsn 430

1459 GGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCAC 1508
     |||||||||||||||||||||||:::|||   ||||||:::::|||
 431 GlyGlyGlyAlaCysLeuPheAsnLysProLeuLysLeuLeuAspProPr 447

1509 GGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTT 1558
     ||||||||||||||:::||||||:::|||||||||||||||||||||
 447 oGluCysGlyAsnGlyPheValGluThrGlyGluGluCysAspCysGlyT 464

1559 TTCATGTGGAATGCTAT.........GGATTATGCTGTAAGAAATGTTCC 1599
        :::||||||                |||||||||||||||:::
 464 hrIleAlaGluCysAlaMetGluGlyGluGluCysCysLysLysCysThr 480

1600 CTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCCC 1649
     |||::::::   ::::::||||||||||||   ||||||:::|||
 481 LeuThrGlnAspSerGluCysSerAspGlyLeuCysCysSerAsn..... 495

1650 ATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGT 1699
       |||   |||:::|||:::         ||||||:::|||||||||:::|
 496 .CysLysPheAsnProLysGluMetLeuCysArgGluAlaValAsnAspC 512

1700 GTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAAT 1749
     |||||||   |||   |||||||||:::::::::|||||||   |||
 512 ysAspIleProGluThrCysThrGlyAsnThrSerGlnCysProAlaAsn 528

1750 CTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTA 1799
     :::|||||   |||||||||:::|||:::::   ||||||   |||::
 529 IleHisLysLeuAspGlyTyrSerCysGluSerMetGlnGlyLeuCysPh 545

1800 CAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAA 1849
     ::::|||:::|||||||||||||||:::||||||:::||||||||||||
 545 eGlyGlyArgCysLysThrArgAspArgGlnCysLysTyrIleTrpGlyG 562

1850 CAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAA 1899
       |||:::::::::::|||::::::|||||||||||||||||   |||
 562 luLysValSerAlaAlaAspArgTyrCysTyrGluLysLeuAsnIleGlu 578

1900 GGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTG 1949
     ||||||||||||||||||||||||::::::   :::   ||||||||||
 579 GlyThrGluLysGlyAsnCysGlyArgAsnLysGluThrTrpIleGlnCy 595

1950 CAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTC 1999
     |:::|||:::||||||:::||||||:::|||||||||||||||:::::::
 595 sAsnLysGlnAspValLeuCysGlyTyrLeuLeuCysThrAsnIleSerA 612

2000 GAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTC 2049
     :::::||||||:::|||:::|||:::|||:::::::      :::|||:::
 612 snValProArgLeuGlyGluLeuAspGlyAspValThrSerSerIle 628

2050 TACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGA 2099
        :::||||||::::::   :::||||||||||:::||||||   |||||
 629 ValAsnGlnGlyLysLeuTyrAsnCysSerGlyGlyHisValLysLeuAs 645

2100 TGATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGT 2149
     |:::||||||||||:::|||||||||||||||||||||||||||||   :
  645 pGluAspThrAspLeuGlyTyrValGluAspGlyThrProCysGlyThrG 662

2150 CTATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATG 2199
     ::||||||||||||:::::::::||||||   |||::::::::::|||:::
  662 lyMetMetCysLeuGluHisArgCysLeuProIleAspSerPheAsnPhe 678

2200 AGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGT 2249
     |||:::|||   |||   |||   :::|||:::||||||||||||||||
  679 SerThrCys...LeuGlySerThrAsnLysIleCysSerGlyHisGlyVa 694

2250 GTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAG 2299
     ||||||||||||:::   ||||||||||      |||:::|||   |
  694 lCysSerAsnGluValArgCysIleCysAspArgPheTrpThrGlyGluA 711

2300 ATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAAGGATGAA 2349
     ||||||||                  ::::::   |||      |||:::
  711 spCysSerSerTyrLeuHisTyrAspHisIleLysProGluGlyAspAsn 727

2350 GGACCCAAGGGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGG 2399
            :::|||      :::||||||:::||||||||:::|||||||
  728 ArgAspGluGlyValIleSerThrAsnIleIleIleGlyAlaIleAlaGl 744

2400 TGCCATCCTGGTAGCAGCTATTGTCCTTGGGGGCACAGGCTGGGGATTTA 2449
     |:::|||||||||   |||:::|||||||   |||:::|||||:::|
  744 yThrIleLeuValLeuAlaLeuValLeuGlyIleThrAlaTrpGlyTyrL 761

2450 AAAATGTCAAGAAGAGAAGGTTCGATCCT 2478
     |||||   :::::::::|||      |||
  761 ysAsnTyrArgArgGluArgGlnIlePro 770
```

seq_name: sp_human:Q14809

seq_documentation_block:

ID Q14809 PRELIMINARY; PRT; 664 AA.
AC Q14809;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE METALLOPROTEASE/DISINTEGRIN-LIKE PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN, OVARY, TESTIS, AND BREAST;
RX MEDLINE=95044425; PubMed=7956356;
RA Katagiri T., Harada Y., Emi M., Nakamura Y.;
RT "Human metalloprotease/disintegrin-like (MDC) gene: exon-intron
RT organization and alternative splicing.";
RL Cytogenet. Cell Genet. 68:39-44(1995).
DR EMBL; D31872; BAA06670.1; -.
DR HSSP; P18619; 1FVL.
DR InterPro; IPR000561; -.
DR InterPro; IPR001590; -.
DR InterPro; IPR001762; -.
DR InterPro; IPR002870; -.
DR Pfam; PF00200; disintegrin; 1.
DR Pfam; PF01421; Reprolysin; 1.
DR Pfam; PF01562; Pep_M12B_propep; 1.
DR PRINTS; PR00289; DISINTEGRIN.
DR PROSITE; PS50215; ADAM_MEPRO; 1.
DR PROSITE; PS50214; DISINTEGRIN_2; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR SMART; SM00050; DISIN; 1.
KW Integrin; Protease; Metalloprotease.
FT NON_TER 1 1
SQ SEQUENCE 664 AA; 72158 MW; 697D074B1106331B CRC64;

alignment_scores:

```
        Quality: 1868.50              Length:    685
          Ratio:  3.664                Gaps:     13
Percent Similarity:  74.453   Percent Identity: 52.117

alignment_block:
US-09-634-252A-2 x Q14809   ..

Align seg 1/1  to: Q14809  from: 1  to: 664

    514 TTGTTGTCTTTCTGATTATGTGGAGATTCACTAC...GAAAATGGGAAACC 560
        ||||||||||||:::|||||||||   |||:::   ::::::|||
      2 LeuLeuSerSerGlnTyrValGluArgHisPheSerArgGluGlyThrTh 18

    561 ACAGTACTCTAAGGGT...GGAGAGCACTGTTACTACCATGGAAGCATCA 607
         |||:::|||   |||   |||:::|||||||||||||:::|||::::::|
     18 rGlnHisSerThrGlyAlaGlyAspHisCysTyrTyrGlnGlyLysLeuA 35

    608 GAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCAT 657
        |||||         |||   :::|||||||||||||||:::|||||||||
     35 rgGlyAsnProHisSerPheAlaAlaLeuSerThrCysGlnGlyLeuHis 51

    658 GGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCT 707
        |||:::|||:::|||   ::::::::::|||:::::|||||   |||::
     52 GlyValPheSerAspGlyAsnLeuThrTyrIleValGluProGlnGluVa 68

    708 GGTT......CATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGA 751
        ::::                  ::         ||||||:::|||   :
     68 lAlaGlyProTrpGlyAlaProGlnGlyProLeuProHisLeuIleTyrA 85

    752 AAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAA 801
        ::|||
     85 rgThr..................................................... 86

    802 AGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAA..... 846
                      |||:::|||   :::            :::
     87 ..............ProLeuLeuProAspProLeuGlyCysArgGluPr 98

    846 ..................................................... 846

     98 oGlyCysLeuPheAlaValProAlaGlnSerAlaProProAsnArgProA 115

    847 .....AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT.........ATA 882
              ||||||||||||   |||:::      ||||||          :::
    115 rgLeuArgArgLysArgGlnValArg......ArgGlyHisProThrVal 129

    883 TTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAAC 932
           :::|||   |||||||:::||||||:::::::::|||||||||:::
    130 HisSerGluThrLysTyrValGluLeuIleValIleAsnAspHisGlnLe 146

    933 GTATAAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGT 982
         :::::::::   |||:::|||   :::   |||:::|||||||||||||
    146 uPheGluGlnMetArgGlnSerValValLeuThrSerAsnPheAlaLysS 163

    983 CCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGG 1032
        ||||||||||||||:::|||   |||||||||||||||||||||||||||
    163 erValValAsnLeuAlaAspValIleTyrLysGluGlnLeuAsnThrArg 179

   1033 GTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACAT 1082
        :::|||||||||||||:::||||||||:::::   |||:::|||:::::
    180 IleValLeuValAlaMetGluThrTrpAlaAspGlyAspLysIleGlnVa 196

   1083 CACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGC 1132
        :      :::   ::::::   |||   ::::::      ||||||::::
    196 lGlnAspAspLeuLeuGluThrLeuAlaArgLeuMetValTyrArgArgG 213

   1133 GC...ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACA 1179
        ::   :::   :::   :::||||||:::||||||:::|||      |||
    213 luGlyLeuProGluProSerAspAlaThrHisLeuPheSerGlyArgThr 229

   1180 TTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCG 1229
        |||:::         |||:::   :::|||   ||||||:::||||||
    230 PheGlnSerThrSerSerGlyAlaAlaTyrValGlyGlyIleCysSerLe 246

   1230 CACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCAC 1279
         :::::::|||   ||||||||||||||||||         |||:::|||
    246 uSerHisGlyGlyGlyValAsnGluTyrGlyAsnMetGlyAlaMetAlaV 263

   1280 AAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGG...GAA 1326
          :::|||:::|||:::|||:::|||||||||||||:::::|||   :::
    263 alThrLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMetMetTrpAsnLys 279

   1327 CCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTG 1376
              |||             |||   |||   :::   |||   |||||
    280 HisArgSerSerAlaGlyAspCysLysCysProAspIleTrpLeuGlyCy 296

   1377 CATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCA 1426
        ||||||||||:::||||||           ||||||||||||:::|||
    296 sIleMetGluAspThrGlyPheTyrLeuProArgLysPheSerArgCysS 313

   1427 GCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTT 1476
        |||||   ||||||:::::|||||||||:::|||||||||:::||||||
    313 erIleAspGluTyrAsnGlnPheLeuGlnGluGlyGlyGlySerCysLeu 329

   1477 TTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATA 1526
        ||||||:::|||   ||||||:::::::|||   ||||||||||||||::
    330 PheAsnLysProLeuLysLeuLeuAspProProGluCysGlyAsnGlyPh 346

   1527 CGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTAT. 1575
        :|||||||||||||||||||||||||||||||||         ||||||
    346 eValGluAlaGlyGluGluCysAspCysGlySerValGlnGluCysSerA 363

   1576 ........GGATTATGCTGTAAGAAATGTTCCCTCTCCAACGGGGCTCAC 1617
                ||||   |||||||||||||:::|||::::::   |||
    363 rgAlaGlyGlyAsnCysCysLysLysCysThrLeuThrHisAspAlaMet 379

   1618 TGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACG 1667
        ||||||||||||   ||||||::::::       |||   ::::::|||||
    380 CysSerAspGlyLeuCysCysArgArg......CysLysTyrGluProAr 394

   1668 AGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATT 1717
        ||||   :::|||||||:::|||||||||||||||||||||:::|||   |
    394 gGlyValSerCysArgGluAlaValAsnGluCysAspIleAlaGluThrC 411

   1718 GTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGA 1767
        ||||||||||||||:::|||||||||||||||||||||||||   ||||||
    411 ysThrGlyAspSerSerGlnCysProProAsnLeuHisLysLeuAspGly 427

   1768 TATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGAC 1817
        |||   |||:::::::::|||||||||||||||:::|||:::||||||||
    428 TyrTyrCysAspHisGluGlnGlyArgCysTyrGlyGlyArgCysLysTh 444

   1818 CAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTG 1867
        |||||||:::|||||||||   :::||||||      ||||||      |
    444 rArgAspArgGlnCysGlnValLeuTrpGlyHisAlaAlaAla......A 459

   1868 ACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAAC 1917
        ||:::|||||||||||||||||||||:::|||||||||||||:::|||:::
    459 spArgPheCysTyrGluLysLeuAsnValGluGlyThrGluArgGlySer 475

   1918 TGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTT 1967
        ||||||:::   |||:::   |||:::|||||||||||||:::||||||::
    476 CysGlyArgLysGlySerGlyTrpValGlnCysSerLysGlnAspValLe 492

   1968 CTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTC 2017
        :|||||||||||||||||:::|||::::::   ||||||||||:::|||:
    492 uCysGlyPheLeuLeuCysValAsnIleSerGlyAlaProArgLeuGlyA 509

   2018 AACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTG 2067
        ::|||   |||:::|||      ::::::|||||||||||||||:::
    509 spLeuValGlyAspIleSerSerValThrPheTyrHisGlnGlyLysGlu 525

   2068 ATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACGGATGTGGG 2117
        :::||||||   |||:::||||||   |||   |||   :::|||:::::
    526 LeuAspCysArgGlyGlyHisValGlnLeuAlaAspGlySerAspLeuSe 542

   2118 CTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATC 2167
```

```
      :|||||||||||||||||||   |||||||||:::|||:::|||||||||:
    542 rTyrValGluAspGlyThrAlaCysGlyProAsnMetLeuCysLeuAspH 559

   2168 GGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGAT 2217
        :::::||||||      :::|||:::|||:::|||:::||||||   :::
    559 isArgCysLeuProAlaSerAlaPheAsnPheSerThrCysProGlySer 575

   2218 TCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCAC 2267
        ::::::   ::::::||||||   |||||||||||||||||||||:::
    576 GlyGluArgArgIleCysSerHisHisGlyValCysSerAsnGluGlyLy 592

   2268 CTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATC 2317
         |||||||||:::      |||:::|||   ||||||||||||:::::|
    592 sCysIleCysGlnProAspTrpThrGlyLysAspCysSerIleHisAsnP 609

    318 CAGTTAGGAACCTTCACCCCCCC...AAGGATGAAGGACCCAAGGGTCCT 2364
        ||:::   :::      |||      :::   |||      |||||||||
    609 roLeuProThrSerProProThrGlyGluThrGluArgTyrLysGlyPro 625

   2365 AGTGCCACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTA3C 2414
        |||:::||||||:::|||||||||||||||||||||||||:::|||||||
    626 SerGlyThrAsnIleIleIleGlySerIleAlaGlyAlaValLeuValAl 642

   2415 AGCTATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAAATGTCAAGAA3A 2464
        ||||||||||||||||||||||||||||||||||||:::::::::
    642 aAlaIleValLeuGlyGlyThrGlyTrpGlyPheLysAsnIleArgArgG 659

   2465 GAAGG 2469
          |||
    659 lyArg 660

seq_name: sp_vertebrate:Q9PSZ3

seq_documentation_block:
ID   Q9PSZ3      PRELIMINARY;      PRT;   452 AA.
AC   Q9PSZ3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   METALLOPROTEASE-DISINTEGRIN (FRAGMENT).
GN   MDC11A.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cai H., Kraetzschmar J., Alfandari D., Hunnicutt G., Blobel C.P.;
RT   "Neural crest-specific and general expression of distinct
RT   metalloprotease-disintegrins in early Xenopus laevis development.";
RL   Dev. Biol. 0:0-0(1998).
DR   EMBL; AF032384; AAC61848.1; -.
DR   HSSP; P18619; 1FVL.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   PRINTS; PR00289; DISINTEGRIN.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   SMART; SM00050; DISIN; 1.
KW   Integrin; Protease; Metalloprotease.
FT   NON_TER       1       1
SQ   SEQUENCE   452 AA;  48577 MW;  DF1F66C24CCD6847 CRC64;


alignment_scores:
            Quality: 1453.00            Length:    447
              Ratio:  4.151              Gaps:      6
 Percent Similarity:  78.300   Percent Identity: 58.166
```

```
alignment_block:
 US-09-634-252A-2 x Q9PSZ3   ..

 Align seg 1/1  to: Q9PSZ3  from: 1  to: 452

   1156 GTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAG 1205
        |||   ||||||:::      ||||||      :::::::|||:::   ::
      9 ValSerLeuIleGlyGlyGlnThrPheLysSerSerHisSerGlyAlaAl 25

   1206 TTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGT 1255
        :|||||||||||:::||||||   |||:::|||   ||||||||||||
     25 aTyrPheGlyGlyIleCysSerProThrHisGlyGlyGlyValAsnGluT 42

   1256 ATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAA 1305
        |||||         |||:::|||   :::|||:::|||:::|||:::|||
     42 yrGlyAsnIleGlyAlaMetAlaIleThrLeuAlaGlnThrLeuGlyGln 58

   1306 AACCTTGGAATCCAATGG...GAACCTTCTAGCAGAAAGCCAAAATGTGA 1352
        |||||||||:::::|||   :::|||   :::              |||
     59 AsnLeuGlyMetMetTrpAsnLysProArgThrThrThrGlyAspCysLy 75

   1353 CTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCC 1402
         |||   :::   |||   |||||||||||||||:::||||||
     75 sCysProAspLeuTrpArgGlyCysIleMetGluAspThrGlyPheTyrL 92

   1403 ATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTA 1452
             :::|||||||||:::||||||:::   ||||||      ||||||
     92 euProGlnLysPheSerArgCysSerValAspGluTyrSerLysPheLeu 108

   1453 CAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGA 1502
        |||   |||||||||:::|||||||||||:::|||   ||||||:::::
    109 GlnAspGlyGlyGlySerCysLeuPheAsnLysProLeuLysLeuLeuAs 125

   1503 GCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATT 1552
        :|||   :::||||||||||||||:::::|||   ||||||||||||||||
    125 pProProSerCysGlyAsnGlyPheIleGluIleGlyGluGluCysAspC 142

   1553 GTGGTTTTCATGTGGAATGC.............TATGGATTATGCTGTAAG 1590
        |||||      :::||||||                 |||   |||||||||
    142 ysGlySerProAlaGluCysAsnLysSerArgAlaGlyAsnCysCysLys 158

   1591 AAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAA 1640
        ||||||:::||||||:::   |||   ||||||||||||   ||||||::
    159 LysCysThrLeuSerHisAspAlaMetCysSerAspGlyLeuCysCysAr 175

   1641 CAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTG 1690
        :      :::|||   ::::::||||||||      ||||||::::::::
    175 g......GlyCysLysTyrGluProArgGlyThrValCysArgGluSerL 190

   1691 TGAACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGC 1740
        ::|||||||||||||:::   |||   |||   |||||||||:::   |||
    190 euAsnGluCysAspValProGluAlaCysProGlyAspSerSerAlaCys 206

   1741 CCACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGG 1790
        |||   ||||||||||||||||||||||   |||:::::::::|||||
    207 ProAlaAsnLeuHisLysGlnAspGlyTyrPheCysAspAsnGluGlnGl 223

   1791 CCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACA 1840
        |||||||::::::|||:::|||||||||||||||||:::||||||:::   :
    223 yArgCysPheGlyGlyArgCysLysThrArgAspArgGlnCysHisAlaL 240

   1841 TCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTG 1890
        ::||||||      |||      ||||||:::||||||||||||||||||
    240 euTrpGlyArgGlyAla......SerAspArgPheCysTyrGluLysLeu 254

   1891 AATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTG 1940
        |||   |||||||||||||||||||||||||:::|||   :::::::||
    255 AsnIleGluGlyThrGluLysGlyAsnCysGlyArgAspArgGlnAsnTr 271

   1941 GATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCA 1990
        ||||||||||||||:::||||||:::||||||:::|||||||||:::|
    271 pIleGlnCysSerLysGlnAspValLeuCysGlyTyrLeuLeuCysSerA 288
```

```
 1991 ATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCA 2040
       ||:::::      |||:::|||||||:::|||:::|||:::|||
  288 snIleSerGlyIleProGlnIleGlyGluLeuAsnGlyAspIleThrSer 304

 2041 ACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGT 2090
         |||||||||||||||:::|||   :::||||||   |||:::::||
  305 MetSerPheTyrHisGlnAsnArgTyrLeuAspCysArgGlyGlyGlnVa 321

 2091 AGTTTTAGATGATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCAT 2140
       |:::|||   |||   :::   :::|||||||||||||||||||||||
  321 lThrLeuProAspGlySerCysLeuGlyTyrValGluAspGlyThrProC 338

 2141 GTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAGCC 2190
       ||||||||:::|||:::|||||||||||:::||||||      :::|||
  338 ysGlyProAsnMetIleCysLeuAspArgArgCysLeuProAlaSerAla 354

 2191 CTAAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGG 2240
       :::|||:::|||:::||||||   :::   :::|||   :::||||||
  355 PheAsnPheSerThrCysProGlySerTrpAsnGlyValIleCysSerAs 371

  241 CCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGG 2290
       |||||||||||||||||||:::   |||||||||         |||:
  371 pHisGlyValCysSerAsnGluGlyLysCysIleCysHisProGluTrpT 388

 2291 CAGGGACAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCC 2340
       ::|||   |||||||||:::   ||||||         |||      |||
  388 hrGlyLysAspCysSerValTyrAspPro.........LeuProValPro 401

 2341 AAGGATGAAGGACCC............AAGGGTCCTAGTGCCACCAATCT 2378
       |||      |||               ||||||||||:::||||||::
  402 LysProThrGlyValValGluLysTyrLysGlyProSerGlyThrAsnIl 418

 2379 CATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCTATTGTCCTTG 2428
       :|||||||||||||||||||||||:::|||:::|||:::|||||||||||||||
  418 eIleIleGlySerIleAlaGlyAlaValLeuIleAlaAlaIleValLeuG 435

 2429 GGGGCACAGGCTGGGGATTTAAAAATGTCAAGAAGAGAAGG 2469
       |||||||||||||||||||||||||||:::::::::   |||
  435 lyGlyThrGlyTrpGlyPheLysAsnIleArgArgGlyArg 448

seq_name: sp_human:Q14810

seq_documentation_block:
ID   Q14810      PRELIMINARY;      PRT;   518 AA.
AC   Q14810;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   METALLOPROTEASE/DISINTEGRIN-LIKE PROTEIN (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BRAIN, OVARY, TESTIS, AND BREAST;
RX   MEDLINE=95044425; PubMed=7956356;
RA   Katagiri T., Harada Y., Emi M., Nakamura Y.;
RT   "Human metalloprotease/disintegrin-like (MDC) gene: exon-intron
RT   organization and alternative splicing.";
RL   Cytogenet. Cell Genet. 68:39-44(1995).
DR   EMBL; D31872; BAA06671.1; -.
DR   HSSP; P18619; 1FVL.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PRINTS; PR00289; DISINTEGRIN.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   SMART; SM00050; DISIN; 1.
KW   Integrin; Protease; Metalloprotease.
FT   NON_TER       1      1
SQ   SEQUENCE   518 AA;  57144 MW;  FB98EFEA03B2254E CRC64;


alignment_scores:
            Quality: 1298.00              Length:    513
              Ratio:  3.546                Gaps:     12
 Percent Similarity:  71.345   Percent Identity: 49.903

alignment_block:
 US-09-634-252A-2 x Q14810   ..

 Align seg 1/1  to: Q14810  from: 1  to: 518

  514 TTGTTGTCTTCTGATTATGTGGAGATTCACTAC...GAAAATGGGAAACC 560
       ||||||||||||:::|||||||||   |||:::   :::::::|||
    2 LeuLeuSerSerGlnTyrValGluArgHisPheSerArgGluGlyThrTh 18

  561 ACAGTACTCTAAGGGT...GGAGAGCACTGTTACTACCATGGAAGCATCA 607
       |||:::|||   |||   |||:::|||||||||||||:::|||::::::|
   18 rGlnHisSerThrGlyAlaGlyAspHisCysTyrTyrGlnGlyLysLeuA 35

  608 GAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCAT 657
       |||||          |||   :::|||||||||||||:::|||||||||
   35 rgGlyAsnProHisSerPheAlaAlaLeuSerThrCysGlnGlyLeuHis 51

  658 GGCATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCT 707
       |||:::|||:::|||   :::::::::|||::::::|||||   |||::
   52 GlyValPheSerAspGlyAsnLeuThrTyrIleValGluProGlnGluVa 68

  708 GGTT......CATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGA 751
       ::::              :::              ||||||:::|||   :
   68 lAlaGlyProTrpGlyAlaProGlnGlyProLeuProHisLeuIleTyrA 85

  752 AAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAA 801
       ::|||
   85 rgThr..................................................... 86

  802 AGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAA..... 846
                    |||:::|||   :::            :::
   87 ..............ProLeuLeuProAspProLeuGlyCysArgGluPr 98

  846 ..................................................... 846

   98 oGlyCysLeuPheAlaValProAlaGlnSerAlaProProAsnArgProA 115

  847 .....AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT.........ATA 882
            ||||||||||   |||:::      ||||||          :::
  115 rgLeuArgArgLysArgGlnValArg......ArgGlyHisProThrVal 129

  883 TTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAAC 932
         :::|||   ||||||:::||||||:::::::::|||||||||:::
  130 HisSerGluThrLysTyrValGluLeuIleValIleAsnAspHisGlnLe 146

  933 GTATAAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGT 982
       :::::::::   |||:::|||   :::   |||:::||||||||||||
  146 uPheGluGlnMetArgGlnSerValValLeuThrSerAsnPheAlaLysS 163

  983 CCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGG 1032
       |||||||||||||:::|||   |||||||||||||||||||||||||||
  163 erValValAsnLeuAlaAspValIleTyrLysGluGlnLeuAsnThrArg 179

 1033 GTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACAT 1082
       :::|||||||||||:::|||||||||::::::   |||:::|||:::::
  180 IleValLeuValAlaMetGluThrTrpAlaAspGlyAspLysIleGlnVa 196

 1083 CACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGC 1132
       :      :::   ::::::   |||   ::::::      ||||||::::
  196 lGlnAspAspLeuLeuGluThrLeuAlaArgLeuMetValTyrArgArgG 213

 1133 GC...ATTAAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACA 1179
```

```
      ::   :::   :::   :::||||||:::||||||:::|||      |||
  213 luGlyLeuProGluProSerAspAlaThrHisLeuPheSerGlyArgThr 229

 1180 TTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCG 1229
      |||:::         |||:::   :::|||   ||||||:::||||||
  230 PheGlnSerThrSerSerGlyAlaAlaTyrValGlyGlyIleCysSerLe 246

 1230 CACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCAC 1279
       ::::::|||   ||||||||||||||||||         |||:::|||
  246 uSerHisGlyGlyGlyValAsnGluTyrGlyAsnMetGlyAlaMetAlaV 263

 1280 AAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGG...GAA 1326
        :::|||:::|||:::|||:::||||||||||||:::::|||   :::
  263 alThrLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMetMetTrpAsnLys 279

  327 CCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTG 1376
          |||          |||   |||   :::   |||   |||||
  280 HisArgSerSerAlaGlyAspCysLysCysProAspIleTrpLeuGlyCy 296

 1377 CATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCA 1426
      ||||||||||:::||||||           ||||||||||||:::||||
  296 sIleMetGluAspThrGlyPheTyrLeuProArgLysPheSerArgCysS 313

 1427 GCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTT 1476
      |||||   ||||||:::::|||||||||:::|||||||||:::||||||
  313 erIleAspGluTyrAsnGlnPheLeuGlnGluGlyGlyGlySerCysLeu 329

 1477 TTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATA 1526
      ||||||:::|||   ||||||::::::|||   ||||||||||||||::
  330 PheAsnLysProLeuLysLeuLeuAspProProGluCysGlyAsnGlyPh 346

 1527 CGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTAT. 1575
      :|||||||||||||||||||||||||||||         ||||||
  346 eValGluAlaGlyGluGluCysAspCysGlySerValGlnGluCysSerA 363

 1576 ........GGATTATGCTGTAAGAAATGTTCCCTCTCCAACGGGGCTCAC 1617
              |||   |||||||||||||||:::|||:::::::   |||
  363 rgAlaGlyGlyAsnCysCysLysLysCysThrLeuThrHisAspAlaMet 379

 1618 TGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACG 1667
      ||||||||||||   ||||||::::::       |||   ::::::|||||
  380 CysSerAspGlyLeuCysCysArgArg......CysLysTyrGluProAr 394

 1668 AGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATT 1717
      ||||   :::||||||:::|||||||||||||||||||||:::|||   |
  394 gGlyValSerCysArgGluAlaValAsnGluCysAspIleAlaGluThrC 411

  718 GTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGA 1767
      ||||||||||||||:::|||||||||||||||||||||||   ||||||
  411 ysThrGlyAspSerSerGlnCysProProAsnLeuHisLysLeuAspGly 427

 1768 TATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGAC 1817
      |||   |||:::::::::|||||||||||||||:::|||:::|||||||
  428 TyrTyrCysAspHisGluGlnGlyArgCysTyrGlyGlyArgCysLysTh 444

 1818 CAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTG 1867
      |||||||:::|||||||||   :::||||||      ||||||      |
  444 rArgAspArgGlnCysGlnValLeuTrpGlyHisAlaAlaAla......A 459

 1868 ACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAAC 1917
      ||:::|||||||||||||||||||||:::||||||||||||:::|||:::
  459 spArgPheCysTyrGluLysLeuAsnValGluGlyThrGluArgGlySer 475

 1918 TGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAA 1956
      ||||||:::   |||:::   |||:::||||||||||||
  476 CysGlyArgLysGlySerGlyTrpValGlnCysSerLys 488

seq_name: sp_human:Q14808

seq_documentation_block:
ID   Q14808      PRELIMINARY;      PRT;   524 AA.
AC   Q14808;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MDC PROTEIN.
GN   MDC.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=CEREBELLUM;
RX   MEDLINE=94073190; PubMed=8252040;
RA   Emi M., Katagiri T., Harada Y., saito H., Inazawa J., Ito I.,
RA   Kasumi F., Nakamura Y.;
RT   "A novel metalloprotease/disintegrin-like gene at 17q21.3 is
RT   somatically rearranged in two primary breast cancers.";
RL   Nat. Genet. 5:151-157(1993).
DR   EMBL; D17390; BAA04213.1; -.
DR   HSSP; P18619; 1FVL.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   ProDom; PD000664; -; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   SMART; SM00050; DISIN; 1.
SQ   SEQUENCE   524 AA;  57901 MW;  5C863E558D906238 CRC64;


alignment_scores:
            Quality: 1297.00              Length:    517
              Ratio:  3.515                Gaps:     12
 Percent Similarity:  71.373   Percent Identity: 49.516

alignment_block:
 US-09-634-252A-2 x Q14808   ..

 Align seg 1/1  to: Q14808  from: 1  to: 524

     502 CTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTAC...GA 548
         |||:::::::   ||||||||||||:::||||||||||   |||:::   ::
       4 LeuSerHisGlnLeuLeuSerSerGlnTyrValGluArgHisPheSerAr 20

     549 AAATGGGAAACCACAGTACTCTAAGGGT...GGAGAGCACTGTTACTACC 595
         ::::|||      |||:::|||   |||   |||:::||||||||||||:
      20 gGluGlyThrThrGlnHisSerThrGlyAlaGlyAspHisCysTyrTyrG 37

     596 ATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGC 645
         ::|||:::::|||||          |||   :::|||||||||||||||
      37 lnGlyLysLeuArgGlyAsnProHisSerPheAlaAlaLeuSerThrCys 53

     646 AATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTGTATATGATAGA 695
         :::|||||||||||:::|||:::|||   :::::::::|||::::::||
      54 GlnGlyLeuHisGlyValPheSerAspGlyAsnLeuThrTyrIleValGl 70

     696 GCCACTAGAGCTGGTT......CATGATGAGAAAAGCACAGGTCGACCAC 739
         ||||   |||::::::              :::          ||||
      70 uProGlnGluValAlaGlyProTrpGlyAlaProGlnGlyProLeuProH 87

     740 ATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAAT 789
         ||:::|||   :::|||
      87 isLeuIleTyrArgThr..................................  92

     790 CTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTG 839
                                     |||:::|||   :::
      93 ..........................ProLeuLeuProAspProLeuGl 100

     840 GTTGAAA.....................................  846
             :::
     100 yCysArgGluProGlyCysLeuPheAlaValProAlaGlnSerAlaProP 117
```

```
 847 .................AGAAGGAAGAGAGCAGTGAATCCATCACGTGGT 879
                      ||||||||||||   |||:::      ||||||
 117 roAsnArgProArgLeuArgArgLysArgGlnValArg......ArgGly 131

 880 .........ATATTTGAAGAAATGAAATATTTGGAACTTATGATTGTTAA 920
             :::   :::|||   ||||||:::||||||::::::::||
 132 HisProThrValHisSerGluThrLysTyrValGluLeuIleValIleAs 148

 921 TGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAACA 970
     |||||||:::   :::::::::   |||:::|||   :::   |||:::|
 148 nAspHisGlnLeuPheGluGlnMetArgGlnSerValValLeuThrSerA 165

 971 ACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAG 1020
     ||||||||||||||||||||||||||:::|||   ||||||||||||||
 165 snPheAlaLysSerValValAsnLeuAlaAspValIleTyrLysGluGln 181

1021 CTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGA 1070
     ||||||||||||:::|||||||||||||:::|||||||||::::::   ||
 182 LeuAsnThrArgIleValLeuValAlaMetGluThrTrpAlaAspGlyAs 198

1071 TCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAA 1120
     |:::|||::::::      :::   ::::::   |||   ::::::
 198 pLysIleGlnValGlnAspAspLeuLeuGluThrLeuAlaArgLeuMetV 215

1121 AATACCGGCAGCGC...ATTAAGCAGCATGCTGATGCTGTGCACCTCATC 1167
       ||||||:::::   :::   :::   ::::::|||:::||||||:::
 215 alTyrArgArgGluGlyLeuProGluProSerAsnAlaThrHisLeuPhe 231

1168 TCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGG 1217
     |||      ||||||:::         |||:::   :::|||   |||||
 232 SerGlyArgThrPheGlnSerThrSerSerGlyAlaAlaTyrValGlyGl 248

1218 TGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAA 1267
     |:::||||||   :::::::|||   ||||||||||||||||||
 248 yIleCysSerLeuSerHisGlyGlyGlyValAsnGluTyrGlyAsnMetG 265

1268 TGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATC 1317
       |||:::|||   :::|||:::|||:::|||:::|||||||||||:::
 265 lyAlaMetAlaValThrLeuAlaGlnThrLeuGlyGlnAsnLeuGlyMet 281

1318 CAATGG...GAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATC 1364
     :::|||   :::      |||              |||   |||   :::
 282 MetTrpAsnLysHisArgSerSerAlaGlyAspCysLysCysProAspIl 298

1365 CTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCATTCTCGAAAAT 1414
       |||   |||||||||||||||:::||||||         |||||||
 298 eTrpLeuGlyCysIleMetGluAspThrGlyPheTyrLeuProArgLysP 315

1415 TTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGT 1464
     |||||:::|||||||||   ||||||:::::::|||||||||:::||||||
 315 heSerArgCysSerIleAspGluTyrAsnGlnPheLeuGlnGluGlyGly 331

1465 GGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACGGAATG 1514
     |||:::|||||||||||||:::|||   ||||||:::::|||   |||||
 332 GlySerCysLeuPheAsnLysProLeuLysLeuLeuAspProProGluCy 348

1515 TGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATG 1564
     ||||||||||:::||||||||||||||||||||||||||||||||
 348 sGlyAsnGlyPheValGluAlaGlyGluGluCysAspCysGlySerValG 365

1565 TGGAATGCTAT.........GGATTATGCTGTAAGAAATGTTCCCTCTCC 1605
       ||||||            |||  ||||||||||||||:::|||:::
 365 lnGluCysSerArgAlaGlyGlyAsnCysCysLysLysCysThrLeuThr 381

1606 AACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCT 1655
     :::   |||   ||||||||||||   ||||||::::::      |||
 382 HisAspAlaMetCysSerAspGlyLeuCysCysArgArg......CysLy 396

1656 TTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATA 1705
      ::::::|||||||||   :::|||||||:::||||||||||||||||||
 396 sTyrGluProArgGlyValSerCysArgGluAlaValAsnGluCysAspI 413

1706 TTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCAT 1755
     ||:::|||   ||||||||||||||:::|||||||||||||||||||
 413 leAlaGluThrCysThrGlyAspSerSerGlnCysProProAsnLeuHis 429

1756 AAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGG 1805
     |||   ||||||||   |||:::::::::|||||||||||||||:::||
 430 LysLeuAspGlyTyrTyrCysAspHisGluGlnGlyArgCysTyrGlyGl 446

1806 CGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGG 1855
     |:::|||||||||||||||:::|||||||||   :::||||||      |
 446 yArgCysLysThrArgAspArgGlnCysGlnValLeuTrpGlyHisAlaA 463

1856 CTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACT 1905
     |||||      |||:::|||||||||||||||||||||:::|||||||||
 463 laAla......AspArgPheCysTyrGluLysLeuAsnValGluGlyThr 477

1906 GAGAAGGGAAACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAA 1955
     |||:::|||:::||||||:::   |||:::   |||:::||||||||||
 478 GluArgGlySerCysGlyArgLysGlySerGlyTrpValGlnCysSerLy 494

1956 A 1956
     |
 494 s 494
```

seq_name: sp_human:Q9H013

seq_documentation_block:

```
ID   Q9H013      PRELIMINARY;      PRT;   918 AA.
AC   Q9H013;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MELTRIN-BETA/ADAM 19 HOMOLOGUE.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BLOOD;
RA   Fritsche J., Andreesen R., Kreutz M.;
RT   "Molecular cloning and characterization of a human metalloprotease
RT   disintegrin - a novel marker for dendritic cell differentiation.";
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Y13786; CAC20585.1; -.
SQ   SEQUENCE   918 AA;  100877 MW;  7FB862870498C2C7 CRC64;
```

alignment_scores:

```
            Quality: 1011.50            Length:    806
              Ratio:  2.094               Gaps:     30
 Percent Similarity:  59.926   Percent Identity: 31.638
```

alignment_block:

US-09-634-252A-2 x Q9H013 ..

Align seg 1/1 to: Q9H013 from: 1 to: 918

```
  97 GTGCCTGCCAGCGCCCCGGCCCGCACGCCGCCCTGCCGCCTGCTTCTCGT 146
     :::|||:::||||||   |||            :::||||||   |||::
   1 MetProGlySerAlaGlyAla............AlaArgLeuCysLeuLe 13

 147 CCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCGCGCCTGGG 196
     :   :::   |||   ||||||           |||||||||
  13 uAlaPheAlaLeuGlnProLeu............ArgProArg....... 23

 197 GGGCTGCTGCGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAAAAAAAT 246
               |||         |||   |||::::::   :::|||::::::
  24 ........AlaAlaArgGluProGlyTrpThrArgGlySerGluGluGly 37

 247 TTGGGAGTCCTGGCAGATGAAGACAATACATTGCAACAGAATAGCAGCAG 296

  37 ................................................. 37
```

```
 297 TAATATCAGTTACAGCAATGCAATGCAGAAAGAAATCACACTGCCTTCAA 346
                 |||      :::|||   |||:::   :::|||:::
  38 ............SerProLysLeuGlnHisGluLeuIleIleProGlnT 50

 347 GACTCATATATTACATCAACCAAGACTCGGAAAGCCCTTATCACGTTCTT 396
                      :::   ||||||||||||
  50 rp.................LysThrSerGluSerProVal......... 57

 397 GACACAAAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCTGGC 446
                                    :::      |||
  58 ...............................ArgGluLysHisProLe 63

 447 CCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACC 496
      :::|||:::::::::::::   |||   |||   :::::::||||||||||
  63 uLysAlaGluLeuArgValMetAlaGluGlyArgGluLeuIleLeuAspL 80

 497 TCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTAC 546
     ||      |||:::   |||::::::   :::|||:::|||   ||||||
  80 euGluLysAsnGluGlnLeuPheAlaProSerTyrThrGluThrHisTyr 96

 547 ...GAAAATGGGAAACCACAGTACTCT......AAGGGTGGAGAGCACTG 587
        :::::|||:::||||||   :::      |||      :::||||||
  97 ThrSerSerGlyAsnProGlnThrThrThrArgLysLeuGluAspHisCy 113

 588 TTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGT 637
     |:::|||||||||:::::::|||   ::::::   |||:::|||:::||||
 113 sPheTyrHisGlyThrValArgGluThrGluLeuSerSerValThrLeuS 130

 638 CAACCTGCAATGGACTTCATGGCATGTTTGAA...GATGATACCTTCGTG 684
     ||||||||:::|||:::::::|||::::::      ::::::::::::
 130 erThrCysArgGlyIleArgGlyLeuIleThrValSerSerAsnLeuSer 146

 685 TATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACAGGTCG 734
     |||:::||||||||||||             |||:::|||:::
 147 TyrValIleGluProLeuPro.........AspSerLysGly........ 157

 735 ACCACATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGA 784
         |||:::|||   :::                    |||:::::::::|
 158 .GlnHisLeuIleTyrArg..................SerGluHisLeuL 168

 785 AG............AATCTCACTATGGAAAGAGGTGAC............ 810
     ||            |||      :::|||::::::
 168 ysProProProGlyAsnCysGlyPheGluHisSerLysProThrThrArg 184

 811 CAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAAGAGAGC 860
     :::|||      |||:::      |||      |||:::|||   |||
 185 AspTrpAla...LeuGlnPheThrGlnGlnThrLysLysArgProArgAr 200

 861 AGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTTA 910
      :::      :::|||      :::::::::|||||||||:::||||||
 200 gMet......LysArgGluAspLeuAsnSerMetLysTyrValGluLeuT 215

 911 TGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCA 960
       :::|||   |||:::      :::::::|||:::|||   ::::::
 215 yrLeuValAlaAspTyrLeuGluPheGlnLysAsnArgArgAspGlnAsp 231

 961 CATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTA 1010
        |||:::::::         :::::::::|||   ||||||:::::::||
 232 AlaThrLysHisLysLeuIleGluIleAlaAsnTyrValAspLysPheTy 248

1011 CAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGA 1060
     |:::   :::||||||   |||:::::::||||||:::::::|||:::||||
 248 rArg...SerLeuAsnIleArgIleAlaLeuValGlyLeuGluValTrpT 264

1061 CTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCAT 1110
     ||:::   ::::::   :::::::::   ||||||   :::   |||
 264 hrHisGlyAsnMetCysGluValSerGluAsnProTyrSerThrLeuTrp 280

1111 GAGTTCTCAAAATACCGGCAGCGCATT......AAGCAGCATGCTGATGC 1154
     :::|||   :::::::|||:::::::::      :::::::::   |||
 281 SerPheLeuSerTrpArgArgLysLeuLeuAlaGlnLysTyrHisAspAs 297

1155 TGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGA 1204
      :::::::||||||:::   :::::::||||||          :::|||:
  297 nAlaGlnLeuIleThrGlyMetSerPheHisGlyThrThrIleGlyLeuA 314

1205 GTTACTTTGGAGGTGTCTGTTCT......CGCACAAGAGGAGTTGGTGTG 1248
     ::   :::   ::::::||||||      ::::::   ||||||::::::
  314 laProLeuMetAlaMetCysSerValTyrGlnSerGlyGlyValAsnMet 330

1249 AATGAGTATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCT 1298
     ::::::           ::::::||||||   ::::::::::::::::
  331 AspHisSerGluAsnAlaIleGlyValAlaAlaThrMetAlaHisGluMe 347

1299 GGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAAT 1348
     :::::::|||:::|||:::      :::   ::::::                |
  347 tGlyHisAsnPheGlyMetThrHisAspSerAlaAsp.........CysC 361

1349 GTGACTGCACAGAATCCTGGGGTGGCTGCATCATG...GAGGAAACAGGG 1395
     ||:::::::::   :::   ||||||||||||||          ||||||
  361 ysSerAlaSerAlaAlaAspGlyGlyCysIleMetAlaAlaAlaThrGly 377

1396 GTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGA 1445
                :::   |||:::   |||:::      |||
  378 HisProPheProLysValPheAsnGlyCysAsnArgArgGluLeuAspAr 394

1446 CTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCA......A 1489
      :::||||||   |||||||||   ||||||   |||   |||
  394 gTyrLeuGlnSerGlyGlyGlyMetCysLeuSerAsnMetProAspThrA 411

1490 CAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGG 1539
          |||:::          :::||||||||||||||||:::|||   |||
  411 rgMetLeuTyrGlyGlyArgArgCysGlyAsnGlyTyrLeuGluAspGly 427

1540 GAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTATGGATTATGCTGT.. 1587
     ||||||||||||||||||   :::   ||||||   :::   ||||||
  428 GluGluCysAspCysGlyGluGluGluGluCysAsnAsnProCysCysAs 444

1588 ....AAGAAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCT 1633
         :::::::|||:::|||       ||||||:::|||:::   |||   |
  444 nAlaSerAsnCysThrLeuArgProGlyAlaGluCysAlaHisGlySerC 461

1634 GCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGG 1683
     |||||:::      :::|||   :::          |||      ||||||
  461 ysCysHis......GlnCysLysLeuLeuAlaProGlyThrLeuCysArg 475

1684 GATGCTGTGAACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGG 1733
     :::   :::::::::||||||:::   |||:::|||||||||   |||
  476 GluGlnAlaArgGlnCysAspLeuProGluPheCysThrGlyLysSerPr 492

1734 TCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAA 1783
      :::||||||   |||:::::::::::||||||      |||:::   :
  492 oHisCysProThrAsnPheTyrGlnMetAspGlyThrProCysGluGlyG 509

1784 ATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGT 1833
     ::|||:::   ||||||||||||   |||   |||   ::::::||||||
  509 lyGlnAlaTyrCysTyrAsnGlyMetCysLeuThrTyrGlnGluGlnCys 525

1834 CAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGA 1883
     |||   :::||||||      |||      :::      :::|||:::||
  526 GlnGlnLeuTrpGlyProGlyAlaArgProAlaProAspLeuCysPheGl 542

1884 AAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAGGAT...G 1930
     ||||:::|||:::   |||          ||||||||||||||||   :
  542 uLysValAsnValAlaGlyAspThrPheGlyAsnCysGlyLysAspMetA 559

1931 GAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTA 1980
     ::   :::      :::|||:::   :::|||:::   ||||||   :::
  559 snGlyGluHisArgLysCysAsnMetArgAspAlaLysCysGlyLysIle 575

1981 CTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGA 2030
        |||   :::        |||          |||::::::::
  576 GlnCysGlnSerSerGluAlaArgPro.............LeuGluSerAs 588

2031 GATCATTCCA......ACTTCCTTCTACCATCAAGGCCGGGTGATTGACT 2074
```

```
      :   :::|||      |||::::::      :::||||||   |||:::|
    588 nAlaValProIleAspThrThrIleIleMetAsnGlyArgGlnIleGlnC 605

   2075 GCAGTGGTGCCCATGTAGTTTTA..............GATGATGATACG 2109
        ||   |||:::||||||                     |||
    605 ysArgGlyThrHisValTyrArgGlyProGluGluGluGlyAspMetLeu 621

   2110 GATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTG 2159
        |||   |||   |||      ||||||   ||||||   :::   :::||
    622 AspProGlyLeuValMetThrGlyThrLysCysGlyTyrAsnHisIleCy 638

   2160 TTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTC 2209
        |::::::   :::|||   :::   :::   ::::::   :::::::|||
    638 sPheGluGlyGlnCysArgAsnThrSerPhePheGluThrGluGlyCys. 654

   2210 CACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAAT 2259
                     ||||||   |||:::|||||||||||||||:::|||
    655 .............GlyLysLysCysAsnGlyHisGlyValCysAsnAsn 666

   2260 GAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTAT 2309
        :::   :::|||   |||         ||||||         |||:::
    667 AsnGlnAsnCysHisCysLeuProGlyTrpAlaProProPheCysAsnTh 683

   2310 CCGGGATCCAGTTAGGAACCTTCACCCCCCCAAGGATGAAGGACCCAAGG 2359
                  ::::::          |||:::||||||
    683 rProGlyHisGlyGlySerIle.....AspSerGlyProMetP 696

   2360 GTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCTG 2409
          |||:::::::::::::   :::::::::::::::      |||||||||
    696 roProGluSerValGlyProValValAlaGlyValLeuValAlaIleLeu 712

   2410 GTAGCAGCTATTGTCCTT 2427
        |||   |||::::::::::
    713 ValLeuAlaValLeuMet 718

seq_name: sp_rodent:O35674

seq_documentation_block:
ID   O35674      PRELIMINARY;      PRT;   920 AA.
AC   O35674;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MELTRIN BETA.
GN   ADAM19 OR MLTNB.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98129833; PubMed=9461614;
RA   Inoue D., Reid M.S., Lum L., Kratzschmar J., Weskamp G., Myung Y.M.,
RA   Baron R., Blobel C.P.;
RT   "Cloning and initial characterization of mouse meltrin beta and
RT   analysis of the expression of four metalloprotease-disintegrins in
RT   bone cells.";
RL   J. Biol. Chem. 273:4180-4187(1998).
RN   [2]
RP   SEQUENCE OF 429-578 FROM N.A.
RX   MEDLINE=96026308; PubMed=7566181;
RA   Yagami-Hiromasa T., Sato T., Kurisaki T., Kamijo K., Nabeshima Y.,
RA   Fujisawa-Sehara A.;
RT   "A metalloprotease-disintegrin participating in myoblast fusion.";
RL   Nature 377:652-656(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98288128; PubMed=9622634;
RA   Kurisaki T., Masuda A., Osumi N., Nabeshima Y., Fujisawa-Sehara A.;
RT   "Spatially- and Temporally-restricted Expression of Meltrin a and b in
RT   Mouse embryo: Possible Roles of Metalloprotease-disintegrins in
RT   Organogenesis and Neurogenesis.";
RL   Mech. Dev. 73:211-215(1998).
DR   EMBL; AF019887; AAC40037.1; -.
DR   EMBL; D50410; BAA18923.2; -.
DR   HSSP; P18619; 1FVL.
DR   MEROPS; M12.214; -.
DR   MGD; MGI:105377; Adam19.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   ProDom; PD000664; -; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW   Integrin; Protease; Metalloprotease.
SQ   SEQUENCE   920 AA;  100860 MW;  7094FDD4EE547382 CRC64;


alignment_scores:
          Quality:  975.00            Length:    800
            Ratio:  2.092               Gaps:     25
 Percent Similarity:  58.250   Percent Identity: 30.500

alignment_block:
 US-09-634-252A-2 x O35674   ..

 Align seg 1/1  to: O35674  from: 1  to: 920

    130 TGCCGCCTGCTTCTCGTCCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTC 179
        |||   |||   |||:::|||   |||      ||||||||||||   ::
     11 CysLeuLeuAlaLeuAlaLeuGlnLeuHisTrpProLeuAlaAlaCysGl 27

    180 CCGGCCCCGCGCCTGGGGGGCTGCTGCGCCCAGCGCTCCGCATTGGAATG 229
        :                                       |||   |||:::
     27 u.....................................ProGlyTrpThrT 32

    230 AAACTGCAGAAAAAAATTTGGGAGTCCTGGCAGATGAAGACAATACATTG 279
           :::::::::
     32 hrArgGlySerGln..................................... 36

    280 CAACAGAATAGCAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGA 329
                              :::   |||      :::|||   ||
     37 .......................GluGlySerProProLeuGlnHisGl 45

    330 AATCACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGAAA 379
        |:::   :::|||:::                     :::   |||||||
     45 uLeuIleIleProGlnTrp.................ArgThrSerGluS 56

    380 GCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAAT 429
        |||||                                        :::
     56 erPro.....................................Gly 58

    430 AAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTC 479
        ::::::   |||      :::|||:::::::::::::   |||   |||
     59 ArgGlyLysHisProLeuArgAlaGluLeuArgValMetAlaGluGlyAr 75

    480 CAAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATT 529
         ::::::||||||||||||      |||:::   |||::::::      |
     75 gGluLeuIleLeuAspLeuGluLysAsnGluHisLeuPheAlaProAlaT 92

    530 ATGTGGAGATTCACTAC...GAAAATGGGAAACCACAGTACTCT...... 570
        ||:::|||      |||      :::|||:::||||||   |||
     92 yrThrGluThrCysTyrThrAlaSerGlyAsnProGlnThrSerThrLeu 108

    571 AAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGA 620
        |||:::   :::||||||:::|||||||||:::::|||   |||   ::
    109 LysSerGluAspHisCysPheTyrHisGlyThrValArgAspValAspGl 125

    621 CTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATG...TTTG 667
```

```
 125 uSerSerValThrLeuSerThrCysArgGlyIleArgGlyLeuIleIleV 142

 668 AAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATGAT 717
 142 alArgSerAsnLeuSerTyrIleIleGluProVal..............  153

 718 GAGAAAAGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCAGGACA 767
 154 ...ProAsnSerAspSerGlnHisArgIleTyrArgSer.........  165

 768 GTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAGGTGAC.......  810
 166 ..............GluHisLeuThrLeuProProGlyAsnCysGlyP  177

 811 ........................CAGTGGCCCTTTCTCTCTGAATTA  834
 177 heGluHisSerGlyProThrSerLysAspTrpAlaLeuGlnPheThrHis  193

 835 CAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATT  884
 194 GlnThrLysLysGlnProArgArgMet.........LysArgGluAspLe  207

 885 TGAAGAAATGAAATATTTGGAACTTATGATTGTTAATGATCACAAAACGT  934
 207 uHisSerMetLysTyrValGluLeuTyrLeuValAlaAspTyrAlaGluP  224

 935 ATAAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGCAAAGTCC  984
 224 heGlnLysAsnArgHisAspGlnAspAlaThrLysArgLysLeuMetGlu  240

 985 GTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACACCAGGGT 1034
 241 IleAlaAsnTyrValAspLysPheTyrArg...SerLeuAsnIleArgIl  256

1035 TGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGACATCA 1084
 256 eAlaLeuValGlyLeuGluValTrpThrHisGlyAspLysCysGluValS  273

1085 CCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGC 1134
 273 erGluAsnProTyrSerThrLeuTrpSerPheLeuSerTrpArgArgLys  289

1135 ATT.........AAGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGT 1175
 290 LeuLeuAlaGlnLysSerHis...AspAsnAlaGlnLeuIleThrGlyAr  305

1176 GACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGGT 1225
 305 gSerPheGlnGlyThrThrIleGlyLeuAlaProLeuMetAlaMetCysS  322

1226 CT......CGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAAGG 1269
 322 erValTyrGlnSerGlyGlyValSerMetAspHisSerGluAsnAlaIle  338

1270 GCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCA 1319
 339 GlyValAlaSerThrValAlaHisGluIleGlyHisAsnPheGlyMetSe  355

1320 ATGGGAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGG 1369
 355 rHisAspSerAlaHis.........CysCysSerAlaSerAlaAlaAspG  369

1370 GTGGCTGCATCATG...GAGGAAACAGGGGTGTCCCATTCTCGAAAATTT 1416
 369 lyGlyCysIleMetAlaAlaAlaThrGlyHisProPheProLysValPhe  385

1417 TCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGG 1466
 386 SerTrpCysAsnArgLysGluLeuAspArgTyrLeuGlnThrGlyGlyG  402

1467 AGCCTGCCTTTTCAACAGGCCA......ACAAAGCTATTTGAGCCCACGG 1510
```

```
 402 yMetCysLeuSerAsnMetProAspThrArgThrLeuTyrGlyGlyArgA 419

1511 AATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTT 1560
 419 rgCysGlyAsnGlyTyrLeuGluAspGlyGluGluCysAspCysGlyGlu 435

1561 CATGTGGAATGCTATGGATTATGCTGT......AAGAAATGTTCCCTCTC 1604
 436 GluGluGluCysLysAsnProCysCysAsnAlaSerAsnCysThrLeuLy 452

1605 CAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTC 1654
 452 sGluGlyAlaGluCysAlaHisGlySerCysCysHis......GlnCysL 467

1655 TTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGAT 1704
 467 ysLeuValAlaProGlyThrGlnCysArgGluGlnValArgGlnCysAsp 483

1705 ATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCA 1754
 484 LeuProGluPheCysThrGlyLysSerProHisCysProThrAsnTyrTy 500

1755 TAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATG 1804
 500 rGlnMetAspGlyThrProCysGluGlyGlyGlnAlaTyrCysTyrAsnG 517

1805 GCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAG 1854
 517 lyMetCysLeuThrTyrGlnGluGlnCysGlnGlnLeuTrpGlyProGly 533

1855 GCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCAC 1904
 534 AlaArgProAlaLeuAspLeuCysPheGluArgValAsnAlaAlaGlyAs 550

1905 TGAGAAGGGAAACTGCGGGAAGGAT...GGAGACCGGTGGATTCAGTGCA 1951
 550 pThrTyrGlyAsnCysGlyLysGlyLeuAsnGlyGlnTyrArgLysCysS 567

1952 GCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGA 2001
 567 erProArgAspAlaLysCysGlyLysIleGlnCysGlnSerThrGlnAla 583

2002 GCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTA 2051
 584 ArgProLeu......GluSerAsnAlaValSerIleAspThrThrIleTh 598

2052 CCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTA.... 2097
 598 rLeuAsnGlyArgArgIleHisCysArgGlyThrHisValTyrArgGlyP 615

2098 ...................GATGATGATACGGATGTGGGCTATGTAGAA 2127
 615 roGluGluGluGluGlyGluGlyAspMetLeuAspProGlyLeuValMet 631

2128 GATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCT 2177
 632 ThrGlyThrLysCysGlyHisAsnHisIleCysPheGluGlyGlnCysAr 648

2178 ACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCCAAGGGTA 2227
 648 gAsnThrSerPhePheGluThrGluGlyCys..............GlyL 660

2228 AAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGT 2277
 660 ysLysCysAsnGlyHisGlyValCysAsnAsnAsnLysAsnCysHisCys 676

2278 GATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGTTAGGAA 2327
 677 PheProGlyTrpSerProProPheCysAsnThrProGlyAspGlyGlySe 693

2328 CCTTCACCCCCCCAAGGATGAAGGACCCAAGGGTCCTAGTGCCACCAATC 2377
 693 rVal............AspSerGlyProLeuProProLysSerValGlyP 706
```

```
    2378 TCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCTATTGTCCTT 2427
          :::|||:::::::::   :::|||:::::::|||   |||:::::::::
     706 roValIleAlaGlyValPheSerAlaLeuPheValLeuAlaValLeuVal 722

seq_name: sp_vertebrate:O12960

seq_documentation_block:
ID   O12960      PRELIMINARY;      PRT;   914 AA.
AC   O12960;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   ADAM 13.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97223528; PubMed=9070330;
RA   Alfandari D., Wolfsberg T.G., White J.M., Desimone D.W.;
RT   "ADAM 13: a novel ADAM expressed in somitic mesoderm and neural crest
RT   cells during Xenopus laevis development.";
RL   Dev. Biol. 182:314-330(1997).
DR   EMBL; U66003; AAB51194.1; -.
DR   HSSP; P18619; 1FVL.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   ProDom; PD000664; -; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR   SMART; SM00181; EGF; 1.
SQ   SEQUENCE   914 AA;  99749 MW;  87B20F8FAD65D085 CRC64;


alignment_scores:
            Quality:  968.50              Length:    689
              Ratio:  2.196                 Gaps:     24
 Percent Similarity:  64.006   Percent Identity: 32.075

[redacted]nment_block:
[redacted]9-634-252A-2 x O12960   ..

[redacted]n seg 1/1  to: O12960  from: 1  to: 914

     432 GGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCA 481
         |||   |||:::|||   |||::::::    :::::::::   |||   |
      56 GlyThrProAspGly...GlyGluIle.LeuValSerSerGluGlyArgL 71

     482 AATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTAT 531
         ||||||||||   :::      |||:::   |||::::::      |||
      71 ysPheIleLeuLysValGluArgAsnHisLeuLeuPheAlaProGlyTyr 87

     532 GTGGAGATTCACTACGAAAATGGGAAA.............CCACAGTA 566
         :::|||   ||||||   :::|||:::              |||:::::
      88 ThrGluThrHisTyrThrAspGlyGlnMetValThrLeuSerProAsnHi 104

     567 CTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCA 616
         ::::         ||||||||||||||||||||:::::::::::::
     104 sThr.........GluHisCysTyrTyrHisGlyGlnValGluAsnTyrA 118

     617 AAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTT 666
           :::|||:::|||||||||:::|||||:::|||:::   |||::::::

     118 spGluSerSerValAlaLeuThrThrCysSerGlyIleSerGlyLeuIle 134

     667 GAAGATGATACC...TTCGTGTATATGATAGAGCCACTAGAGCTGGTTCA 713
               :::|||         |||   ::::::|||||||||:::
     135 TrpLeuSerThrAsnAsnSerTyrTyrLeuLysProLeuGluValProGl 151

     714 TGATGAGAAAAGCACA......GGTCGACCACATATAATCCAGAAAACCT 757
         |||       |||               |||::::::   |||
     151 yLysGluThrHisThrLeuValArgThrGluHisLeuLeuIleLysGluG 168

     758 TGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAGGT 807
           :::      ::::::                              :::   |||
     168 lySerCysGlyHisAspGly.....................HisSerGly 177

     808 GACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAAGAG 857
         :::         :::|||:::|||:::          :::   :::   ::
     178 SerThrAlaSerTyrLeuGlnGluPheThrAlaProSerSerHisHisHi 194

     858 AGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATTTGGAAC 907
         :   |||:::      |||:::::::::::::::::::|||||||:::||||
     194 sArgValArg......ArgAsnValTrpArgSerGlnLysTyrMetGluL 209

     908 TTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCAT 957
         ||:::||||||   |||::::::   :::   |||:::::::   :::
     209 euPheIleValAlaAspTyrSerMetPheMetLysGlnAsnArgAsnLeu 225

     958 GCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTAT 1007
         :::   |||::::::   :::   :::::::::|||   ||||||:::::
     226 GlySerThrLysGlnArgValLeuGluIleAlaAsnTyrValAspLysPh 242

    1008 TTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCT 1057
         :|||      ::::::|||   :::|||:::|||:::::::::|||:::|
     242 eTyrMet...SerMetAsnIleLysValAlaLeuIleGlyLeuGluValT 258

    1058 GGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAGATGCTC 1107
         ||||||||:::||||||   :::::::::   :::      :::   |||
     258 rpThrGluArgAspGlnCysGluValAsnAspAspAlaAsnAspSerLeu 274

    1108 CATGAGTTCTCAAAATACCGGCAGCGCATT.........AAGCAGCATGC 1148
           :::|||   :::::::::|||:::::::        |||:::|||
     275 LysSerPheLeuGlnTrpLysGlnLysLeuArgSerArgLysLysHis.. 290

    1149 TGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCA 1198
          |||   ::::::||||||:::   |||||||||              :
     291 .AspAsnAlaGlnLeuIleThrGlyValThrPheLysGlyThrThrIleG 307

    1199 GTCTGAGTTACTTTGGAGGTGTCTGTTCT......CGCACAAGAGGAGTT 1242
         ::::::::   :::   |||:::|||:::      ::::::   ||||||
     307 lyMetAlaProLeuGluGlyMetCysThrAlaGluAsnSerGlyGlyVal 323

    1243 GGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCA 1292
         ::::::::::::            :::::::::|||   ::::::::::
     324 SerMetAspHisSerGluAsnAlaIleGlyAlaAlaAlaThrMetAlaHi 340

    1293 GAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGC 1342
         :::::::::::::|||:::|||:::               |||:::
     340 sGluIleGlyHisAsnPheGlyMet..............SerHisAspA 352

    1343 CAAAATGT......GACTGCACAGAATCCTGGGGTGGCTGCATCATG... 1383
              |||      ::::::|||   :::   ||||||||||||||||
     352 spGlyCysCysValGluAlaThrProGluGlnGlyGlyCysIleMetAla 368

    1384 GAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTT 1433
           ||||||          ||||||||||||:::||||||
     369 AlaAlaThrGlyHisProPheProArgLysPheSerSerCysSerGlnLy 385

    1434 GGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACA 1483
          :::      ::::::::::|||:::|||||||||   ||||||||||
     385 sGlnLeuMetSerTyrPheGlnLysGlyGlyGlyMetCysLeuPheAsnM 402

    1484 GGCCA...ACAAAG...CTATTTGAGCCCACGGAATGTGGAAATGGATAC 1527
           |||   ||||||   |||            :::||||||||||||:::
     402 etProAsnThrLysAspLeuValMetGlyLysLysCysGlyAsnGlyPhe 418
```

```
1528 GTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTATGG 1577
     :::|||   ||||||:::|||||||||||||         ||||||   ::
 419 LeuGluGluGlyGluGlnCysAspCysGlyGluProGluGluCysThrAs 435

1578 ATTATGCTGT......AAGAAATGTTCCCTCTCCAACGGGGCTCACTGCCA 1621
     :   ||||||      :::::::|||:::|||:::   |||||||:::|||:
 435 nSerCysCysAsnAlaAsnAsnCysThrLeuLysAlaGlyAlaGlnCysA 452

1622 GCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGACGG 1671
     ::   |||   ||||||::::::      |||   ::::::      |||
 452 laHisGlyGluCysCysGlnAsp......CysLysLeuLysSerAlaGly 466

1672 TATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTAC 1721
        :::|||||||:::   :::::::::|||||||:::   |||:::|||||
 467 ThrGlnCysArgGluMetAlaGlySerCysAspLeuProGluPheCysTh 483

1722 TGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATG 1771
     |||||||:::   :::||||||   |||:::::|||   ||||||
 483 rGlyAspAlaProSerCysProSerAsnValTyrLysLeuAspGlySerL 500

1772 CATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGA 1821
       |||   ::::::::::::   ||||||||||||   |||   |||:::
 500 euCysAlaAspGlyAsnAlaTyrCysTyrAsnGlyMetCysLeuThrHis 516

1822 GACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAA 1871
     ::::::||||||   ::::::||||||:::   |||:::   :::   ::
 517 GlnGlnGlnCysIleHisLeuTrpGlySerGlyAlaValValAlaProAs 533

1872 GTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCG 1921
     :||||||::::::   :::|||      |||   :::   ||||||||||
 533 nPheCysPheGlnAspValAsnLysAlaGlyAspGlnTyrGlyAsnCysG 550

1922 GGAAGGATGGA...GACCGGTGGATTCAGTGCAGCAAACATGATGTGTTC 1968
     |||||:::|||      ::::::::::::|||:::::::::|||:::
 550 lyLysAsnGlyArgGlyGlnPheValLysCysThrSerArgAspAlaLys 566

1969 TGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCA 2018
     ||||||   :::   |||   :::   ::::::   ||||||      ::
 567 CysGlyLysIleGlnCysGlnThrSerSerGluLysProArgAspProSe 583

2019 ACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGA 2068
     ::::            :::   :::::::::      :::|||      :
 583 rMetValLys......ValAspAsnThrIleIleIleAsnGlyTyrLysM 598

2069 TTGACTGCAGTGGTGCCCATGTA......GTTTTAGATGATGATACGGAT 2112
     ::   |||:::|||:::|||:::      :::   :::::::::   |||
 598 etLysCysGlnGlyValHisAlaTyrSerMetGlnGluGluGluGlyAsp 614

 113 GTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTT 2162
        |||   |||      ||||||   ||||||   :::|||:::|||
 615 ProGlyLeuValMetThrGlyThrLysCysGlyAspGlyMetValCysLy 631

2163 AGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCAC 2212
       ||||||:::|||                  |||   |||         |
 631 sAspArgArgCys..............GlnAsnAlaSerPhePheGluL 643

2213 TCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAA 2262
     |||||:::      :::   |||:::|||||||||||||||:::::::::
 643 euAspGlnCysValSerLysCysAsnGlyHisGlyValCysAsnSerAsn 659

2263 GCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTATCCG 2312
        :::|||   ||||||      ||||||          |||
 660 ArgAsnCysHisCysAspSerGlyTrpAlaProProTyrCys........ 673

2313 GGATCCAGTTAGGAACCTTCACCCCCCCCAAGGATGAAGGACCCAAGGGTC 2362
                :::          |||      :::      :::||||
 674 .............AspLysProGlyProGlyGlySerGlnAspSerGlyP 686

2363 CTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTA 2412
     ||:::   :::::::|||   :::|||                :::::::::
 686 roAlaProSerAspLeuProValGlyValThrIlePheLeuValIleLeu 702
```

```
    2413 GCAGCTATTGTCCTT 2427
               :::::::|||
     703 PheLeuValLeuLeu 707
```

```
seq_name: sp_rodent:Q61824

seq_documentation_block:
ID   Q61824      PRELIMINARY;      PRT;   903 AA.
AC   Q61824;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MELTRIN ALPHA.
GN   ADAM12.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96026308; PubMed=7566181;
RA   Yagami-Hiromasa T., Sato T., Kurisaki T., Kamijo K., Nabeshima Y.,
RA   Fujisawa-Sehara A.;
RT   "A metalloprotease-disintegrin participating in myoblast fusion.";
RL   Nature 377:652-656(1995).
DR   EMBL; D50411; BAA08912.1; -.
DR   HSSP; P18619; 1FVL.
DR   MEROPS; M12.212; -.
DR   MGD; MGI:105378; Adam12.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   ProDom; PD000664; -; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR   SMART; SM00050; DISIN; 1.
SQ   SEQUENCE   903 AA;  98600 MW;  6EB0208D2023DCE3 CRC64;


alignment_scores:
          Quality:  966.00              Length:    822
            Ratio:  2.055                Gaps:     32
 Percent Similarity:  57.178   Percent Identity: 31.630

alignment_block:
 US-09-634-252A-2 x Q61824   ..

 Align seg 1/1  to: Q61824  from: 1  to: 903

     103 GCCAGCGCCCCGGCC...CGCACGCCGCCCTGCCGCCTGCTTCTCGTCCT 149
         |||:::   ||||||   |||:::||||||:::|||   |||||||:::
       2 AlaGluArgProAlaArgArgAlaProProAlaArgAlaLeuLeuLeuAl 18

     150 TCTC.........CTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCGCG 190
          |||          ||||||   |||   ||||||   :::
      18 aLeuAlaGlyAlaLeuLeuAlaProArgAlaAlaArgGlyMet...SerL 34

     191 CCTGGGGGGCTGCTGCGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAA 240
           |||
      34 euTrp..................................................... 35

     241 AAAAATTTGGGAGTCCTGGCAGATGAAGACAATACATTGCAACAGAATAG 290
                                                 :::|||:::::
      36 .....................................AspGlnArgGl 39

     291 CAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAATCACACTGC 340
         ::::   :::::::::   ::::::   ::::::|||:::      :::|
      39 yAlaTyrGluValAlaArgAlaSerLeuLeuSerLysAspProGlyIleP 56
```

```
 341 CTTCAAGACTCATATATTACATCAACCAAGACTCGGAAAGCCCTTATCAC 390
     ||:::::   |||
  56 roGlyGlnSerIle..................................... 60

 391 GTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCA 440
               |||:::                                ||
  61 .............ProAlaLysAsp........................Hi 65

 441 TCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTC 490
     |            :::   |||:::::      :::      :::||||
  65 sProAspValLeuThrValGlnLeuGlnLeuGluSerArgAspLeuIleL 82

 491 TTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATT 540
     ||:::|||      |||:::||||||:::::::::   ::::::|||
  82 euSerLeuGluArgAsnGluGlyLeuIleAlaAsnGlyPheThrGluThr 98

 541 CACTACGAAAATGGGAAACCACAGTACTCT.........AAGGGTGGAGA 581
     ||||||   :::         :::   |||         :::      ::
  99 HisTyrLeuGlnAspGlyThrAspValSerLeuThrArgAsnHisThrAs 115

  82 GCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGG 631
     :|||||||||||||||||   ::::::|||         |||   |||:
 115 pHisCysTyrTyrHisGlyHisValGlnGlyAspAlaAlaSerValValS 132

 632 CTCTGTCAACCTGCAATGGACTTCATGGC......ATGTTTGAAGATGAT 675
     ::|||||||||||:::   |||:::|||      |||||||||:::
 132 erLeuSerThrCysSerAspLeuArgGlyLeuIleMetPheGluAsnLys 148

 676 ACCTTCGTGTATATGATAGAGCCACTA.....................GA 704
     |||      |||   :::||||||:::                     ::
 149 Thr......TyrSerLeuGluProMetLysAsnThrThrAspSerTyrLy 163

 705 GCTGGTTCATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAA 754
     :||||||      |||:::   |||:::          ||||||
 163 sLeuValProAlaGluSerMetThrAsn............IleGlnGlyL 176

 755 CCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGA 804
         ::::::|||::::::|||      :::||||||||||||||||
 176 euCysGlySerGlnHisAsnLys......SerAsnLeuThrMetGlu... 189

 805 GGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAA 854
     |||      |||     |||::::::
 190 ...AspValSerProGlyThrSerGlnMet..................... 198

 855 GAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATTTGG 904
      ||||||   :::   :::|||      ::::::      |||||||:|
 199 .ArgAlaArgArgHisLysArgGluThrLeuLysMetThrLysTyrValG 215

 905 AACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCT 954
     |||||:::||||||   |||:::::::   ::::::::::::   ::::::
 215 luLeuValIleValAlaAspAsnArgGluPheGlnArgGlnGlyLysAsp 231

 955 CATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTC 1004
             :::::::::         :::::::::|||   ||||||::
 232 LeuGluLysValLysGlnArgLeuIleGluIleAlaAsnHisValAspLy 248

1005 TATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGA 1054
     ::::|||:::      ||||||   |||:::||||||||||:::||||||:
 248 sPheTyrArg...ProLeuAsnIleArgIleValLeuValGlyValGluV 264

1055 CCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAGATG 1104
     ::|||::::::   |||:::   :::|||:::   :::|||
 264 alTrpAsnAspIleAspLysCysSerIleSerGlnAspProPheThrArg 280

1105 CTCCATGAGTTCTCAAAATACCGGCAGCGCATTAAG............CA 1142
     ||||||||||||      :::|||   :::||||||         ::
 281 LeuHisGluPheLeuAspTrpArg...LysIleLysLeuLeuProArgLy 296

1143 GCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGA 1192
     :      |||   :::::||||||||||   |||   |||:::
 296 sSerHisAspAsnAlaGlnLeuIleSerGlyValTyrPheGlnGlyThrT 313

1193 GAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCT......CGCACAAGA 1236
          :::::::::   :::   ::::::|||:::      ::::::
  313 hrIleGlyMetAlaProIleMetSerMetCysThrAlaGluGlnSerGly 329

1237 GGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGGCACAAGTATT 1286
     ||||||   :::::::::         |||::::::::|||   :::||
  330 GlyValValMetAspHisSerAspSerProLeuGlyAlaAlaValThrLe 346

1287 ATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCA 1336
     |:::::::::|||::::::|||:::|||::::::   :::      :::|
  346 uAlaHisGluLeuGlyHisAsnPheGlyMetAsnHisAspThrLeuGluA 363

1337 GAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCATGGAG 1386
     ||      :::|||      :::   :::   |||||||||||||||:::
  363 rgGlyCysSerCysArgMetAlaAlaGluLysGlyGlyCysIleMetAsn 379

1387 ...GAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTT 1433
        :::||||||                  ||||||:::||||||
  380 ProSerThrGlyPheProPheProMetValPheSerSerCysSerArgLy 396

1434 GGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACA 1483
      :::   :::      |||:::::::|||   |||   ||||||||||||
  396 sAspLeuGluAlaSerLeuGluLysGlyMetGlyMetCysLeuPheAsnL 413

1484 GGCCA......ACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATAC 1527
       |||         :::   |||         :::||||||||||||||||
  413 euProGluValLysGlnAlaPheGlyGlyArgLysCysGlyAsnGlyTyr 429

1528 GTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCTATGG 1577
     ||||||   ||||||||||||||||||||         ||||||   ::
  430 ValGluGluGlyGluGluCysAspCysGlyGluProGluGluCysThrAs 446

1578 ATTATGCTGTAAG......AAATGTTCCCTCTCCAACGGGGCTCACTGCA 1621
     :   |||||||:::         |||:::|||:::      |||   |||:
  446 nArgCysCysAsnAlaThrThrCysThrLeuLysProAspAlaValCysA 463

1622 GCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGG 1671
     ::   |||   ||||||::::::      |||   ::::::|||   |||
  463 laHisGlyGlnCysCysGluAsp......CysGlnLeuLysProProGly 477

1672 TATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTAC 1721
          ||||||   :::   |||:::||||||:::   |||:::|||||
  478 ThrAlaCysArgGlySerSerAsnSerCysAspLeuProGluPheCysTh 494

1722 TGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATG 1771
     ||||   :::   :::||||||   |||::::::   :::||||||:::
  494 rGlyThrAlaProHisCysProAlaAsnValTyrLeuHisAspGlyHisP 511

1772 CATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGA 1821
       |||:::      :::|||   ||||||||||   |||:::|||:::
  511 roCysGlnGlyValAspGlyTyrCysTyrAsnGlyIleCysGlnThrHis 527

1822 GACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAA 1871
     ::::::||||||      :::||||||      |||      :::
  528 GluGlnGlnCysValThrLeuTrpGlyProGlyAlaLysProAlaProGl 544

1872 GTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCG 1921
       :::|||:::|||::::::|||:::   |||         ||||||||||
  544 yIleCysPheGluArgValAsnSerAlaGlyAspProTyrGlyAsnCysG 561

1922 GGAAGGATGGA...GACCGGTGGATTCAGTGCAGCAAACATGATGTGTTC 1968
     ||||||||:::   :::   :::   :::|||:::   :::|||:::
  561 lyLysAspSerLysSerAlaPheAlaLysCysGluLeuArgAspAlaLys 577

1969 TGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCA 2018
     ||||||   :::   |||   :::   :::      |||   ||||||
  578 CysGlyLysIleGlnCysGlnGlyGlyAlaSerArgProValIleGly.. 593

2019 ACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAA......GGCC 2062
          ::::::      |||   |||::::::   :::|||      ||||
  594 .ThrAsnAlaValSerIleGluThrAsnIleProGlnGlnGluGlyGlyA 610

2063 GGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGAT...ACG 2109
```

```
     ||::::::   |||   |||:::||||||   |||   ||||||
 610 rgIleLeu...CysArgGlyThrHisValTyrLeuGlyAspAspMetPro 625

2110 GATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTG 2159
     |||   |||   |||       ||||||   |||:::   :::   :::||
 626 AspProGlyLeuValLeuAlaGlyThrLysCysAlaGluGlyLysIleCy 642

2160 TTTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTC 2209
     ||||:::|||:::|||   :::|||:::::::::::::::::   :::|||
 642 sLeuAsnArgArgCysGlnAsnIleSerValPheGlyValHisLysCysA 659

2210 CACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAAT 2259
       ::::::               |||   |||:::|!!!!!!!!:::|||
 659 laMetGln.............CysHisGlyArgGlyValCysAsnAsn 670

2260 GAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTAT 2309
     :::   :::|||   |||:::      ||||||
 671 ArgLysAsnCysHisCysGluAlaHisTrpAla..................  681

2310 CCGGGATCCAGTTAGGAACCTTCACCCCCCC..................... 2340
                              ||||||
  682 .........................ProProPheCysAspLysPheGlyP 690

2341 ..........AAGGATGAAGGACCC...AAGGGTCCTAGTGCCACCAAT 2376
                  |||:::||||||   :::      :::      :::
  690 heGlyGlySerThrAspSerGlyProIleArgGlnAlaAspAsnGlnGly 706

2377 CTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCTATTGTCCT 2426
     |||   :::|||                ||||||||:::   :::   ||
  707 LeuThrValGly..............IleLeuValSerIleLeuCysLe 718

2427 TGGGGGCACAGGCTGG 2442
     |   ::::::|||:::
  718 uLeuAlaAlaGlyPhe 723
```

seq_name: sp_human:Q13443

seq_documentation_block:

```
ID   Q13443      PRELIMINARY;      PRT;   819 AA.
AC   Q13443; Q10718;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CELLULAR DISINTEGRIN-RELATED PROTEIN PRECURSOR (EC 3.4.24.-)
DE   (METALLOPROTEASE/DISINTEGRIN/CYSTEINE-RICH PROTEIN 9) (MDC9) (MYELOMA
DE   CELL METALLOPROTEINASE) (MCMP).
GN   ADAM9 OR MDC9 OR KIAA0021.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC   TISSUE=BREAST CARCINOMA;
RX   MEDLINE=96178079; PubMed=8647900;
RA   Weskamp G., Kraetzschmar J., Reid M.S., Blobel C.P.;
RT   "MDC9, a widely expressed cellular disintegrin containing cytoplasmic
RT   SH3 ligand domains.";
RL   J. Cell Biol. 132:717-726(1996).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE=96404892; PubMed=8809033;
RA   McKie N., Dallas D.J., Edwards T., Apperley J.F., Russell R.G.G.,
RA   Croucher P.I.;
RT   "Cloning of a novel membrane-linked metalloproteinase from human
RT   myeloma cells.";
RL   Biochem. J. 318:459-462(1996).
RN   [3]
RP   SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC   TISSUE=MYELOID;
RX   MEDLINE=96051387; PubMed=7584026;
RA   Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayasi Y.,
RA   Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
RT   "Prediction of the coding sequences of unidentified human genes. I.
RT   The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
RT   analysis of randomly sampled cDNA clones from human immature myeloid
RT   cell line KG-1.";
RL   DNA Res. 1:27-35(1994).
RN   [4]
RP   REVISIONS.
RA   Nomura N.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: MAY MEDIATE CELL-CELL OR CELL-MATRIX INTERACTIONS (BY
CC       SIMILARITY).
CC   -!- FUNCTION: PROBABLE ZINC PROTEASE.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC       METALLOPROTEASE); ALSO KNOWN AS THE REPROLYSIN SUBFAMILY.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
DR   EMBL; U41766; AAC50403.1; -.
DR   EMBL; D14665; BAA03499.2; -.
DR   HSSP; P18619; 1FVL.
DR   MIM; 602713; -.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   ProDom; PD000664; -; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR   SMART; SM00181; EGF; 1.
KW   Signal; Glycoprotein; Transmembrane; EGF-like domain; Protease;
KW   Hydrolase; Metalloprotease; Zinc; Metal-binding.
FT   SIGNAL        1     28       POTENTIAL.
FT   CHAIN        29    819       CELLULAR DISINTEGRIN-RELATED PROTEIN.
FT   DOMAIN       29    697       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    698    718       POTENTIAL.
FT   DOMAIN      719    819       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      413    503       DISINTEGRIN-LIKE.
FT   DOMAIN      505    634       CYS-RICH.
FT   DOMAIN      644    698       EGF-LIKE.
FT   DOMAIN      790    795       POLY-PRO.
FT   METAL       347    347       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    348    348       BY SIMILARITY.
FT   METAL       351    351       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       357    357       ZINC (CATALYTIC) (BY SIMILARITY).
FT   CARBOHYD    125    125       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    144    144       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    154    154       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    231    231       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    381    381       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    487    487       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT      1    118       MISSING (IN REF. 2).
FT   CONFLICT    117    117       R -> Q (IN REF. 4).
FT   CONFLICT    119    135       YVEGVHNSSIALSDCFG -> MWREFIIHPLLLATVLD
FT                                (IN REF. 2).
FT   CONFLICT    154    154       N -> M (IN REF. 2).
FT   CONFLICT    566    566       G -> GLSLKFHAPFLSTMLQEAVRQTGTYLGGSVCCMKSD
FT                                CRIVTLVK (IN REF. 2).
FT   CONFLICT    713    735       AIFIFIKRDQLWRSYFRKKRSQT -> DYFYLHQEGSTVEK
FT                                LLQKEEITN (IN REF. 2).
FT   CONFLICT    736    819       MISSING (IN REF. 2).
SQ   SEQUENCE   819 AA;  90555 MW;  BC186641833137FF CRC64;
```

alignment_scores:

```
            Quality:  920.50            Length:    766
              Ratio:  2.073               Gaps:     24
 Percent Similarity:  57.963   Percent Identity: 30.809
```

```
alignment_block:
 US-09-634-252A-2 x Q13443   ..

 Align seg 1/1  to: Q13443  from: 1  to: 819

    103 GCCAGCGCCCCGGCCCGCACGCCGCCCTGCCGCCTGCTTCTCGTCCTTCT 152
        |||      |||:::   |||         |||   ||||||:::|||
      5 AlaArgPheProSerGlyThrLeuArgValArgTrpLeuLeuLeuGl 21

    153 CCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCGCGCCTGGGGGGCTG 202
         |||:::   |||:::   ::::::::::||||||
     21 yLeuValGlyProValLeuGlyAlaAlaArgProGlyPhe..... 34

    203 CTGCGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAAAAAAATTTGGGA 252

     34 ................................................. 34

    253 GTCCTGGCAGATGAAGACAATACATTGCAACAGAATAGCAGCAGTAATAT 302
                                    |||::::::|||
     35 ............................GlnGlnThrSerHisLeuSe 41

     03 CAGTTACAGCAATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCA 352
         ||||||                |||||      |||   ||||||
     41 rSerTyr..................GluIleIleThrProTrpArgLeuT 52

    353 TATATTACATCAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACA 402
                   ::::::::::::      ||||||
     52 hrArgGlu...ArgArgGluAlaProArgProTyrSer..... 63

    403 AAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGC 452
                                                  |||::
     64 ........................................LysGlnVa 66

    453 AAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATAC 502
        :|||:::   |||:::|||   |||::::::   |||:::   |||
     66 lSerTyrValIleGlnAlaGluGlyLysGluHisIleIleHisLeuGluA 83

    503 TGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAAAT 552
          |||:::   ||||||   :::|||:::|||
     83 rgAsnLysAspLeuLeuProGluAspPheValValTyr..... 95

    553 GGGAAACCACAGTACTCTAAGGGTGGA..................... 579
                   |||:::|||   |||
     96 .........ThrTyrAsnLysGluGlyThrLeuIleThrAspHisProAs 109

    580 .......GAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACT 622
             :::||||||:::|||:::|||   ::::::||||||   :::|
    109 nIleGlnAsnHisCysHisTyrArgGlyTyrValGluGlyValHisAsnS 126

    623 CCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGAT 672
        ||::::::|||||||||   |||   ||||||:::|||:::::::::
    126 erSerIleAlaLeuSerAspCysPheGlyLeuArgGlyLeuLeuHisLeu 142

    673 GATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAA 722
        ::::::      |||   ||||||||||||                :::::
    143 GluAsnAlaSerTyrGlyIleGluProLeu..............GlnAs 154

    723 AAGCACAGGTCGACCACATATAATC.....................C 748
        :|||:::         |||||||||                    :
    154 nSerSerHisPheGluHisIleIleTyrArgMetAspAspValTyrLysG 171

    749 AGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATG 798
        ::         :::|||      :::|||::::::::::::   |||
    171 luProLeuLysCysGlyValSerAsnLysAspIleGluLysGluThrAla 187

    799 GAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAG 848
        :::      ::::::   |||         ::::::      |||   ||
    188 LysAspGluGluGluGluPro......ProSerMetThrGlnLeuLeuAr 202

    849 AAGGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAAT 898
        ||||:::||||||||||   |||                  :::   :::|
    202 gArgArgArgAlaValLeuPro.................GlnThrArgT 213

    899 ATTTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGC 948
        ||:::||||||:::||||||   |||   :::   |||
    213 yrValGluLeuPheIleValValAspLysGluArgTyrAspMetMetGly 229

    949 TCTTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGT 998
           :::::::::   :::::::::             ::::::|||   ::
    230 ArgAsnGlnThrAlaValArgGluGluMetIleLeuLeuAlaAsnTyrLe 246

    999 GGATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTG 1048
        :||||||:::|||      :::||||||   |||:::|||||||||:::::
    246 uAspSerMetTyr...IleMetLeuAsnIleArgIleValLeuValGlyL 262

   1049 TAGAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTG 1098
        ::|||   ||||||:::   :::   |||:::|||:::   :::
    262 euGluIleTrpThrAsnGlyAsnLeuIleAsnIleValGlyGlyAlaGly 278

   1099 CAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATT.........AA 1139
        ::::::|||   :::|||   ::::::|||::::::::::         ::
    279 AspValLeuGlyAsnPheValGlnTrpArgGluLysPheLeuIleThrAr 295

   1140 GCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATA 1189
        :::::::   |||::::::::::|||:::   :::       |||
    295 gArgArgHisAspSerAlaGlnLeuValLeuLysLysGlyPheGly...G 311

   1190 AGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGA 1239
              :::::::::::::::   |||   ||||||||||||:::::::::
    311 lyThrAlaGlyMetAlaPheValGlyThrValCysSerArgSerHisAla 327

   1240 GTTGGTGTGAATGAGTATGGT......CTTCCAATGGCAGTGGCACAAGT 1283
           |||:::|||   :::|||                :::   |||:::::
    328 GlyGlyIleAsnValPheGlyGlnIleThrValGluThrPheAlaSerIl 344

   1284 ATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTA 1333
        :::::::::::::|||:::::::||||||||:::::::   :::   :::
    344 eValAlaHisGluLeuGlyHisAsnLeuGlyMetAsnHisAspAspGlyA 361

   1334 GCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCATG 1383
                         ||||||:::   :::      :::|||||||||
    361 rg..............AspCysSerCysGlyAlaLysSerCysIleMet 372

   1384 GAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTT 1433
           ::::::|||:::|||   ||||||:::||||||:::||||||
    373 ...AsnSerGlyAlaSerGlySerArgAsnPheSerSerCysSerAlaGl 388

   1434 GGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAAC. 1482
          :::::::::   :::      :::   ||||||   ||||||:::|||
    388 uAspPheGluLysLeuThrLeuAsnLysGlyGlyAsnCysLeuLeuAsnI 405

   1483 .....AGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATAC 1527
             :::|||   :::   ::::::      :::||||||||||
    405 leProLysProAspGluAlaTyrSerAlaProSerCysGlyAsnLysLeu 421

   1528 GTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGC...TA 1574
        |||:::|||||||||||||||||||||||||         ||||||
    422 ValAspAlaGlyGluGluCysAspCysGlyThrProLysGluCysGluLe 438

   1575 TGGATTATGCTGTAAG......AAATGTTCCCTCTCCAACGGGGCTCACT 1618
              ||||||:::         |||:::|||:::::   |||:::|
    438 uAspProCysCysGluGlySerThrCysLysLeuLysSerPheAlaGluC 455

   1619 GCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGA 1668
        ||:::   |||   ||||||:::::       |||   |||   |||
    455 ysAlaTyrGlyAspCysCysLysAsp......CysArgPheLeuProGly 469

   1669 GGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTG 1718
        |||      ||||||      ::::::||||||||:::   ||||||||
    470 GlyThrLeuCysArgGlyLysThrSerGluCysAspValProGluTyrCy 486

   1719 TACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGAT 1768
        |:::|||:::|||      |||   |||::::::      |||:::||||
    486 sAsnGlySerSerGlnPheCysGlnProAspValPheIleGlnAsnGlyT 503

   1769 ATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACC 1818
```

```
     ||   |||::::::|||::::::   ||||||||||||   |||:::
  503 yrProCysGlnAsnAsnLysAlaTyrCysTyrAsnGlyMetCysGlnTyr 519

 1819 AGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGA 1868
        |||   |||||||||   |||:::|||:::||||||   ::::::
  520 TyrAspAlaGlnCysGlnValIlePheGlySerLysAlaLysAlaAlaPr 536

 1869 CAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACT 1918
       |||   |||:::   :::::::|||:::::::|||   :::   |||||||
  536 oLysAspCysPheIleGluValAsnSerLysGlyAspArgPheGlyAsnC 553

 1919 GCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTC 1968
      |||||   :::|||:::::::::   :::|||:::      :::::::::
  553 ysGlyPheSerGlyAsnGluTyrLysLysCysAlaThrGlyAsnAlaLeu 569

 1969 TGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCA 2018
      ||||||   |||   |||   |||:::   :::   |||   :::|||
  570 CysGlyLysLeuGlnCysGluAsnValGlnGluIleProValPheGly.. 585

 2019 ACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGA 2068
                 |||:::|||:::   :::          :::|||
  586 .............IleValProAlaIleIleGlnThrProSerArgGlyT 598

 2069 TTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACG...GATGTG 2115
            |||   |||:::          |||   :::|||:::   |||
  598 hrLysCysTrpGlyValAspPheGlnLeuGlySerAspValProAspPro 614

 2116 GGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGA 2165
      |||   |||::::::||||||   ||||||   :::   :::|||   ::
  615 GlyMetValAsnGluGlyThrLysCysGlyAlaGlyLysIleCysArgAs 631

 2166 TCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCG 2215
      :   :::|||::::::   ::::::||||||      :::|||   ::::
  631 nPheGlnCysValAspAlaSerValLeuAsnTyr...AspCysAspValG 647

 2216 ATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCC 2265
      :::::|||         |||   ||||||||||||||||:::::::::
  647 lnLysLys.........CysHisGlyHisGlyValCysAsnSerAsnLys 660

 2266 ACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTATCCGG 2313
      :::|||   |||:::      ||||||      :::|||:::   :::
  661 AsnCysHisCysGluAsnGlyTrpAlaProProAsnCysGluThrLys 676

seq_name: sp_human:Q9UHP2

seq_documentation_block:
ID   Q9UHP2      PRELIMINARY;      PRT;   857 AA.
AC   Q9UHP2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   DISINTEGRIN AND METALLOPROTEINASE DOMAIN 19 (FRAGMENT).
GN   ADAM19.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Xu R., Cai J., Ying B., Wang F., Xu T., Zhao S., Li C.;
RT   "Partial sequence of Homo sapiens ADAM19.";
RL   Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF134707; AAF22162.1; -.
DR   HSSP; P18619; 1FVL.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PRINTS; PR00289; DISINTEGRIN.
DR   ProDom; PD000664; -; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR   SMART; SM00050; DISIN; 1.
KW   Integrin.
FT   NON_TER       1      1
SQ   SEQUENCE   857 AA;  94007 MW;  DB93FAD7FFBD495B CRC64;


alignment_scores:
            Quality:  916.00              Length:    655
              Ratio:  2.186                 Gaps:     25
 Percent Similarity:  63.969   Percent Identity: 32.824

alignment_block:
 US-09-634-252A-2 x Q9UHP2   ..

 Align seg 1/1  to: Q9UHP2  from: 1  to: 857

  550 AATGGGAAACCACAGTACTCT......AAGGGTGGAGAGCACTGTTACTA 593
      :::|||:::||||||   :::      |||      :::||||||:::||
    1 SerGlyAsnProGlnThrThrThrArgLysLeuGluAspHisCysPheTy 17

  594 CCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCT 643
      |||||||:::::|||   ::::::   |||:::|||:::||||||||||
   17 rHisGlyThrValArgGluThrGluLeuSerSerValThrLeuSerThrC 34

  644 GCAATGGACTTCATGGCATGTTTGAA...GATGATACCTTCGTGTATATG 690
      ||:::|||:::::::|||::::::      :::::::::::   |||:::
   34 ysArgGlyIleArgGlyLeuIleThrValSerSerAsnLeuSerTyrVal 50

  691 ATAGAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACAGGTCGACCACA 740
      ||||||||||||            |||:::|||:::            ||
   51 IleGluProLeuPro.........AspSerLysGly.........GlnHi 61

  741 TATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAG.... 786
      |:::|||   :::                    |||:::::::::|||
   61 sLeuIleTyrArg..................SerGluHisLeuLysProP 72

  787 ........AATCTCACTATGGAAAGAGGTGAC............CAGTGG 816
              |||      :::|||::::::                :::|||
   72 roProGlyAsnCysGlyPheGluHisSerLysProThrThrArgAspTrp 88

  817 CCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAA 866
          |||:::      |||      |||:::|||   |||   :::
   89 Ala...LeuGlnPheThrGlnGlnThrLysLysArgProArgArgMet.. 103

  867 TCCATCACGTGGTATATTTGAAGAAATGAAATATTTGGAACTTATGATTG 916
          :::|||      :::::::::||||||||:::||||||   :::|
  104 ....LysArgGluAspLeuAsnSerMetLysTyrValGluLeuTyrLeuV 119

  917 TTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACC 966
      ||   |||:::      ::::::|||:::|||   ::::::      |||
  119 alAlaAspTyrLeuGluPheGlnLysAsnArgArgAspGlnAspAlaThr 135

  967 AACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGA 1016
      ::::::         ::::::::::|||   ||||||:::::|||:::
  136 LysHisLysLeuIleGluIleAlaAsnTyrValAspLysPheTyrArg.. 151

 1017 GCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGA 1066
       :::||||||   |||:::::||||||:::::::|||:::||||||:::
  152 .SerLeuAsnIleArgIleAlaLeuValGlyLeuGluValTrpThrHisG 168

 1067 AGGATCAGATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTC 1116
        ::::::   :::::::::   ||||||   :::   |||   :::|||
  168 lyAsnMetCysGluValSerGluAsnProTyrSerThrLeuTrpSerPhe 184

 1117 TCAAAATACCGGCAGCGCATT......AAGCAGCATGCTGATGCTGTGCA 1160
         :::::|||:::::::::      :::::::::   |||   :::::
  185 LeuSerTrpArgArgLysLeuLeuAlaGlnLysTyrHisAspAsnAlaGl 201
```

```
1161 CCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACT 1210
     :|||||||:::  ::::::||||||          :::|||:::   :
 201 nLeuIleThrGlyMetSerPheHisGlyThrThrIleGlyLeuAlaProL 218

1211 TTGGAGGTGTCTGTTCT......CGCACAAGAGGAGTTGGTGTGAATGAG 1254
     ::   ::::::||||||      ::::::   ||||||||||||||||
 218 euMetAlaMetCysSerValTyrGlnSerGlyGlyValAsnMetAspHis 234

1255 TATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCA 1304
                :::::||||||   ::::::::::::::::::::::::
  235 SerGluAsnAlaIleGlyValAlaAlaThrMetAlaHisGluMetGlyHi 251

1305 AAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATGTGACT 1354
     :|||:::|||:::      :::   ::::::            |||:::
  251 sAsnPheGlyMetThrHisAspSerAlaAsp.........CysCysSerA 265

1355 GCACAGAATCCTGGGGTGGCTGCATCATG...GAGGAAACAGGGGTGTCC 1401
     :::::   :::   ||||||||||||||||         ||||||
  265 laSerAlaAlaAspGlyGlyCysIleMetAlaAlaAlaThrGlyHisPro 281

1402 CATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTT 1451
           :::   |||:::   |||:::      |||         :::||
  282 PheProLysValPheAsnGlyCysAsnArgArgGluLeuAspArgTyrLe 298

1452 ACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCA......ACAAAGC 1495
     ||||   |||||||||   ||||||   |||   |||              |
  298 uGlnSerGlyGlyGlyMetCysLeuSerAsnMetProAspThrArgMetL 315

1496 TATTTGAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAG 1545
     ||:::         :::||||||||||||||||:::|||   ||||||||
  315 euTyrGlyGlyArgArgCysGlyAsnGlyTyrLeuGluAspGlyGluGlu 331

1546 TGTGATTGTGGTTTTCATGTGGAATGCTATGGATTATGCTGT......AA 1589
     ||||||||||||   :::   ||||||   :::   ||||||      ::
  332 CysAspCysGlyGluGluGluCysAsnAsnProCysCysAsnAlaSe 348

1590 GAAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTA 1639
     ::::|||:::|||      ||||||:::|||:::   |||   ||||||:
  348 rAsnCysThrLeuArgProGlyAlaGluCysAlaHisGlySerCysCysH 365

1640 ACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCT 1689
     ::      :::|||   :::         |||      ||||||:::
  365 is......GlnCysLysLeuLeuAlaProGlyThrLeuCysArgGluGln 379

1690 GTGAACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTG 1739
     :::::::::||||||:::   |||:::|||||||||   |||   :::||
  380 AlaArgGlnCysAspLeuProGluPheCysThrGlyLysSerProHisCy 396

1740 CCCACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGG 1789
     ||||   |||:::::::::::::||||||      |||:::   :::|||:
  396 sProThrAsnPheTyrGlnMetAspGlyThrProCysGluGlyGlyGlnA 413

1790 GCCGCTGCTACAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTAC 1839
     ::   |||||||||||.|   |||   |||   ::::::|||||||||
  413 laTyrCysTyrAsnGlyMetCysLeuThrTyrGlnGluGlnCysGlnGln 429

1840 ATCTGGGGAACAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCT 1889
     :::||||||      |||      :::      :::|||:::||||||::
  430 LeuTrpGlyProGlyAlaArgProAlaProAspLeuCysPheGluLysVa 446

1890 GAATACAGAAGGCACTGAGAAGGGAAACTGCGGGAAG......GATGGAG 1933
     :|||:::   |||         ||||||||||||||      :::|||:
  446 lAsnValAlaGlyAspThrPheGlyAsnCysGlyLysValMetAsnGlyG 463

1934 ACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTC 1983
     :::::      :::|||:::   :::|||:::   ||||||   :::
  463 luHis...ArgLysCysAsnMetArgAspAlaLysCysGlyLysIleGln 478

1984 TGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGAT 2033
     |||   :::          |||           |||:::::::::
  479 CysGlnSerSerGluAlaArgPro.........LeuGluSerAsnAl 491

2034 CATTCCA......ACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGCA 2077
```

```
        :::|||      |||::::::      :::||||||   |||:::|||
  491 aValProIleAspThrThrIleIleMetAsnGlyArgGlnIleGlnCysA 508

2078 GTGGTGCCCATGTAGTTTTA......GATGATGATACGGATGTG...... 2115
       |||:::||||||           :::::::::   |||:::
  508 rgGlyThrHisValTyrArgGlyProGluGluGluGlyAspMetLeuAsn 524

2116 ...GGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTT 2162
        |||   |||      ||||||   ||||||   :::   :::|||::
  525 ProGlyLeuValMetThrGlyThrLysCysGlyTyrAsnHisIleCysPh 541

2163 AGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCAC 2212
     ::::   :::|||   :::   :::   ::::::   :::::::|||
  541 eGluGlyGlnCysArgAsnThrSerPhePheGluThrGluGlyCys.... 556

2213 TCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAA 2262
                ||||||   |||:::||||||||||||||||:::|||:::
  557 ..........GlyLysLysCysAsnGlyHisGlyValCysAsnAsnAsn 569

2263 GCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTATCCG 2312
        :::|||   |||         ||||||         |||:::
  570 GlnAsnCysHisCysLeuProGlyTrpAlaProProPheCysAsnThrPr 586

2313 GGATCCAGTTAGGAACCTTCACCCCCCCAAGGATGAAGGACCCAAGGGTC 2362
                ::::::           |||:::||||||      |
  586 oGlyHisGlyGlySerIle.....AspSerGlyProMetProP 599

2363 CTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTA 2412
     ||:::::::::::   :::::::::::::::      ||||||||||||
  599 roGluSerValGlyProValValAlaGlyValLeuValAlaIleLeuVal 615

2413 GCAGCTATTGTCCTT 2427
        |||:::::::::
  616 LeuAlaValLeuMet 620
```

seq_name: sp_rodent:Q61072

seq_documentation_block:

```
ID   Q61072      PRELIMINARY;      PRT;   845 AA.
AC   Q61072; Q60618; Q61853;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CELLULAR DISINTEGRIN-RELATED PROTEIN PRECURSOR (MELTRIN GAMMA).
GN   ADAM9 OR MLTNG OR MDC9.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LUNG;
RX   MEDLINE=96178079; PubMed=8647900;
RA   Weskamp G., Kraetzschmar J.R., Reid M., Blobel C.P.;
RT   "MDC9, a widely expressed cellular disintegrin containing cytoplasmic
RT   SH3 ligand domains.";
RL   J. Cell Biol. 132:717-726(1996).
RN   [2]
RP   SEQUENCE OF 426-575 FROM N.A.
RX   MEDLINE=96026308; PubMed=7566181;
RA   Yagami-Hiromasa T., Sato T., Kurisaki T., Kamijo K., Nabeshima Y.,
RA   Fujisawa-Sehara A.;
RT   "A metalloprotease-disintegrin participating in myoblast fusion.";
RL   Nature 377:652-656(1995).
RN   [3]
RP   SEQUENCE OF 432-478 FROM N.A.
RC   STRAIN=BALB/C;
RX   MEDLINE=94195820; PubMed=8146185;
RA   Weskamp G., Blobel C.P.;
RT   "A family of cellular proteins related to snake venom disintegrins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 91:2748-2751(1994).
CC   -!- FUNCTION: MAY MEDIATE CELL-CELL OR CELL-MATRIX INTERACTIONS.
CC   -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
```

```
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED.
CC   -!- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
DR   EMBL; U06145; AAA18424.1; -.
DR   EMBL; U41765; AAC52446.1; -.
DR   EMBL; D50412; BAA08913.1; -.
DR   HSSP; P18619; 1FVL.
DR   MEROPS; M12.209; -.
DR   MGD; MGI:105376; Adam9.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   ProDom; PD000664; -; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR   SMART; SM00181; EGF; 1.
KW   Signal; Glycoprotein; Transmembrane; EGF-like domain; Zinc;
KW   Metal-binding.
FT   SIGNAL        1     29       POTENTIAL.
FT   CHAIN        30    845       CELLULAR DISINTEGRIN-RELATED PROTEIN.
FT   DOMAIN       30    698       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    699    718       POTENTIAL.
FT   DOMAIN      719    845       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      413    503       DISINTEGRIN-LIKE.
FT   DOMAIN      505    634       CYS-RICH.
FT   DOMAIN      644    698       EGF-LIKE.
FT   METAL       347    347       ZINC (CATALYTIC) (BY SIMILARITY).
FT   ACT_SITE    348    348       BY SIMILARITY.
FT   METAL       351    351       ZINC (CATALYTIC) (BY SIMILARITY).
FT   METAL       357    357       ZINC (CATALYTIC) (BY SIMILARITY).
FT   CARBOHYD    144    144       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    154    154       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    231    231       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    381    381       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    487    487       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    636    636       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   845 AA;  92049 MW;  17562E89D4A61674 CRC64;


alignment_scores:
          Quality:  892.00            Length:    803
            Ratio:  1.943               Gaps:     26
[redacted]ent Similarity:  57.161   Percent Identity: 29.639

alignment_block:
 US-09-634-252A-2 x Q61072   ..

 Align seg 1/1  to: Q61072  from: 1  to: 845

     103 GCCAGCGCCCCGGCCCGCACGCCGCCCTGCCGCCTGCTTCTCGTCCTTCT 152
         |||   :::|||      :::         |||   |||||||:::
       5 AlaLeuSerProLeuAlaSerLeuArgLeuArgTrpLeuLeuAlaCysGl 21

     153 CCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCGCGCCTGGGGGGCTG 202
         ||||||   |||:::      ::::::||||||
      21 yLeuLeuGlyProValLeuGluAlaGlyArgPro..............  32

     203 CTGCGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAAAAAAATTTGGGA 252

      32 ..................................................  32

     253 GTCCTGGCAGATGAAGACAATACATTGCAACAGAATAGCAGCAGTAATAT 302
                 |||   ::::::|||::::::   :::|||
      33 .........AspLeuGluGlnThrValHisLeuSerSer.........  42

     303 CAGTTACAGCAATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCA 352
                                 ||||||      |||   ||||||

      43 .....................TyrGluIleIleThrProTrpArgLeu.  51

     353 TATATTACATCAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACA 402
                                                       |||
      52 ..................................................Thr  52

     403 AAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCATCTGGCCCAGGC 452
         :::   |||::::::           ::::::            |||
      53 ArgGluArgArgGluAlaLeuGlyProSerSer.........GlnGlnIl  66

     453 AAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATAC 502
          |||:::   |||:::|||   |||::::::   |||:::   |||
      66 eSerTyrValIleGlnAlaGlnGlyLysGlnHisIleIleHisLeuGluA  83

     503 TGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAAAT 552
           |||:::   ||||||   :::|||:::|||         |||:::
      83 rgAsnThrAspLeuLeuProAsnAspPheValValTyrThrTyrAsp...  98

     553 GGGAAACCACAGTACTCTAAGGGTGGA..................... 579
                         |||   |||
      99 ..................LysGluGlySerLeuLeuSerAspHisProAs 109

     580 .......GAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACT 622
                :::||||||:::|||:::|||   ::::::||||||::::::|
     109 nValGlnSerHisCysHisTyrArgGlyTyrValGluGlyValGlnAsnS 126

     623 CCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGAT 672
         ||   ||||||:::|||:::|||   ||||||:::|||:::::::::
     126 erAlaValAlaValSerAlaCysPheGlyLeuArgGlyLeuLeuHisLeu 142

     673 GATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAA 722
         ::::::      :::   ||||||||||||                :::::
     143 GluAsnAlaSerPheGlyIleGluProLeu..............HisAs 154

     723 AAGCACAGGTCGACCACATATAATC.....................CAGA 751
         :|||:::         ||||||:::                      |||:
     154 nSerSerHisPheGluHisIlePheTyrProMetAspGlyIleHisGlnG 171

     752 AAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAA 801
         ::   |||            |||::::::      :::
     171 luProLeuArgCysGlyValSerAsnArgAspThrGluLysGluGlyThr 187

     802 AGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAAG 851
         :::||||||:::            |||      |||   |||:::|||||
     188 GlnGlyAspGluGluGluHisProSerValThrGlnLeuLeuArgArgAr 204

     852 GAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATATT 901
         |                     |||:::::::::   :::   :::|||:
     204 g.....................ArgAlaValLeuProGlnThrArgTyrV 214

     902 TGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCT 951
         ::||||||:::||||||   |||   :::   |||
     214 alGluLeuPheIleValValAspLysGluArgTyrAspMetMetGlyArg 230

     952 TCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGA 1001
         :::::::::   :::::::::            ::::::|||   :::||
     231 AsnGlnThrAlaValArgGluGluMetIleArgLeuAlaAsnTyrLeuAs 247

    1002 TTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTAG 1051
         ||||:::|||      :::||||||   |||:::|||||||||::::::|
     247 pSerMetTyr...IleMetLeuAsnIleArgIleValLeuValGlyLeuG 263

    1052 AGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCAG 1101
         ||   ||||||:::::::::   |||:::|||      :::      :::
     263 luIleTrpThrAspArgAsnProIleAsnIleIleGlyGlyAlaGlyAsp 279

    1102 ATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATT.........AAGCA 1142
         :::|||   :::|||   ::::::|||:::::::::          :::::
     280 ValLeuGlyAsnPheValGlnTrpArgGluLysPheLeuIleThrArgAr 296

    1143 GCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGA 1192
         ::::   |||:::::::::|||:::   :::      |||
     296 gArgHisAspSerAlaGlnLeuValLeuLysLysGlyPheGly...GlyT 312
```

```
1193 GAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTT 1242
      :::::::::::::::   |||   ||||||||||||:::::::::
 312 hrAlaGlyMetAlaPheValGlyThrValCysSerArgSerHisAlaGly 328

1243 GGTGTGAATGAGTATGGT......CTTCCAATGGCAGTGGCACAAGTATT 1286
     |||:::|||   :::|||              :::   |||::::::::
 329 GlyIleAsnValPheGlyGlnIleThrValGluThrPheAlaSerIleVa 345

1287 ATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCA 1336
     ::::::::::|||:::::|||||||||:::::   :::   :::
 345 lAlaHisGluLeuGlyHisAsnLeuGlyMetAsnHisAspAspGlyArg. 361

1337 GAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCATGGAG 1386
            :::|||   |||       :::       :::|||||||||
 362 ........GluCysPheCysGlyAlaLys......SerCysIleMet... 372

1387 GAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGA 1436
     ::::::|||:::|||   ||||||:::||||||:::||||||      ::
 373 AsnSerGlyAlaSerGlySerArgAsnPheSerSerCysSerAlaGluAs 389

1437 GTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAAC.... 1482
     :::::::   :::       :::   |||||||:::|||||||:::|||
 389 pPheGluLysLeuThrLeuAsnLysGlyGlySerCysLeuLeuAsnIleP 406

1483 ..AGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATACGTG 1530
       :::|||   :::   ::::::       :::|||||||||      |||
 406 roLysProAspGluAlaTyrSerAlaProSerCysGlyAsnLysLeuVal 422

1531 GAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGC...TATGG 1577
     :::   ||||||||||||||||||||       ||||||
 423 AspProGlyGluGluCysAspCysGlyThrAlaLysGluCysGluValAs 439

1578 ATTATGCTGTAAG......AAATGTTCCCTCTCCAACGGGGCTCACTGCA 1621
         |||||||:::         |||:::|||::::::   |||:::|||:
 439 pProCysCysGluGlySerThrCysLysLeuLysSerPheAlaGluCysA 456

1622 GCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGG 1671
     ::   |||   ||||||::::::       |||   |||   |||   |||
 456 laTyrGlyAspCysCysLysAsp......CysGlnPheLeuProGlyGly 470

1672 TATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTAC 1721
           ||||||      ::::::||||||||:::   |||||||||::
 471 SerMetCysArgGlyLysThrSerGluCysAspValProGluTyrCysAs 487

1722 TGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATG 1771
     :|||:::|||      |||||||||::::::      |||:::|||||||
 487 nGlySerSerGlnPheCysProProAspValPheIleGlnAsnGlyTyrP 504

1772 CATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGACCAGA 1821
       |||:::::::::::::::   ||||||||||||   |||:::
  04 roCysGlnAsnSerLysAlaTyrCysTyrAsnGlyMetCysGlnTyrTyr 520

  22 GACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACAA 1871
     |||   |||||||||   |||:::|||:::||||||   ::::::   ::
 521 AspAlaGlnCysGlnValIlePheGlySerLysAlaLysAlaAlaProAr 537

1872 GTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGCG 1921
     :   |||:::   ::::::|||:::::|||   :::   ||||||||||
 537 gAspCysPheIleGluValAsnSerLysGlyAspArgPheGlyAsnCysG 554

1922 GGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGT 1971
     ||   :::|||:::::::::   :::|||:::      :::::::::|||
 554 lyPheSerGlySerGluTyrLysLysCysAlaThrGlyAsnAlaLeuCys 570

1972 GGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACT 2021
     |||   |||   |||   |||:::         |||   :::|||
 571 GlyLysLeuGlnCysGluAsnValGlnAspMetProValPheGly..... 585

2022 TCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTGATTG 2071
               |||:::|||:::   :::         :::|||
 586 ..........IleValProAlaIleIleGlnThrProSerArgGlyThrL 599

2072 ACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACG...GATGTGGGC 2118
        |||   |||:::         |||   :::|||:::   |||   |||
 599 ysCysTrpGlyValAspPheGlnLeuGlySerAspValProAspProGly 615

2119 TATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTTTAGATCG 2168
        |||:::::::||||||   |||       :::   :::|||   :::
 616 MetValAsnGluGlyThrLysCysAspAlaGlyLysIleCysArgAsnPh 632

2169 GAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATT 2218
       :::|||::::::   ::::::||||||   :::                |||
 632 eGlnCysValAsnAlaSerValLeuAsnTyrAsp.........CysAspI 646

2219 CCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACC 2268
       :::|||||||   |||   ||||||||||||||||:::::::   :::
 646 leGlnGlyLys...CysHisGlyHisGlyValCysAsnSerAsnLysAsn 661

2269 TGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCC 2318
     |||   |||:::      ||||||         |||:::
 662 CysHisCysGluAspGlyTrpAlaProProHisCysAsp.......... 674

2319 AGTTAGGAACCTTCACCCCCCCAAGGATGAAGGACCCAAG...GGTCCTA 2365
       :::::::::                  |||:::||||||      :::   :
 675 .ThrLysGlyTyrGlyGlySerValAspSerGlyProThrTyrAsnAlaL 691

2366 GTGCCACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCTGGTAGCA 2415
     :::::|||   |||       |||   ::::::      :::|||::::::
 691 ysSerThrAlaLeuArgAspGlyLeuLeuValPhePhePheLeuIleVal 707

2416 GCTATTGTC 2424
        :::|||
 708 ProLeuVal 710
```

```
seq_name: sp_invertebrate:Q9VXL1

seq_documentation_block:
ID   Q9VXL1      PRELIMINARY;      PRT;   685 AA.
AC   Q9VXL1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CG9163 PROTEIN.
GN   MMD OR CG9163.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
```

```
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RT   Science 287:2185-2195(2000).
DR   EMBL; AE003500; AAF48548.1; -.
DR   HSSP; P18619; 1FVL.
DR   FlyBase; FBgn0041109; mmd.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PRINTS; PR00289; DISINTEGRIN.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR   SMART; SM00050; DISIN; 1.
SQ   SEQUENCE   685 AA;  75625 MW;  02AC9C506678C7F3 CRC64;


alignment_scores:
            Quality:  883.00              Length:    685
              Ratio:  2.138                Gaps:     27
 Percent Similarity:  60.292   Percent Identity: 31.241

alignment_block:
 US-09-634-252A-2 x Q9VXL1    ..

 Align seg 1/1  to: Q9VXL1  from: 1  to: 685

     517 TTGTCTTCTGATTATGTGGAGATTCACTACGAAAATGGGAAACCACAGTA 566
         :::::::::   |||   :::|||   |||                :::::
      25 IleGlyGlyHisTyrTyrGlnIleIleTyrProValGlnLeuArgHisHi 41

     567 CTCTAAGGGTGGA.................................G 580
         ::::|||   |||                                 |
      41 sGluLysMetGlyIleSerThrArgGluValSerLeuProLysAspIleG 58

     581 AGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTG 630
         ||||||||||||||||||||:::::::::       :::::::::
      58 luHisCysTyrTyrHisGlyThrValLysAspTyrProGlyAlaSerAla 74

     631 GCTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATGATACCTT 680
         |||:::   ||||||||||||:::   |||::::::::      :::
      75 AlaPheHisThrCysAsnGlyValSerGlyValIleHisIleGlyAsnGl 91

     681 CGTGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACAG 730
          :::::::::|||:::|||:::         :::             :::
      91 uThrPheValIleHisProPhe.........TyrGlyGlyAspLeuSerL 105

     731 GTCGACCACATATAATC......CAGAAAACCTTGGCAGGACAGTATTCT 774
           :::|||||||:::|||         :::|||   |||::::::
     105 ysHisProHisValIlePheGluAlaArgThrLysAlaAsnLysGlyCys 121

     775 AAGCAAATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCT 824
           :::      ||||||              |||:::|||
     122 AlaAsnSerGlyAsnLeu..............AspSerTrpArg..... 131

     825 CTCTGAATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCA. 873
                       |||:::|||||||   :::   ::::::   :::
     132 ..............LeuSerArgArgThrLysHisLeuSerAlaGlyV 143

     874 ..CGTGGTATATTTGAAGAAATG.......................... 894
           |||:::   |||||:::
     143 alAlaGlyValValGluGluIleLeuGlnAlaGlyValGlyArgAsnLys 159

     895 .....................AAATATTTGGAACTTATGATTGTTAATGA 923
                                |||||||:::|||      |||:::   ||
     160 ArgAspValArgGluAlaThrLysTyrIleGluThrAlaIleIleValAs 176

     924 TCACAAAACGTAT...AAGAAGCATCGCTCTTCTCATGCACATACCAACA 970
         |          :::   |||:::::::   |||:::::::|||::::::   :
     176 pLysAlaMetPheAspLysArgAsnGlySerThrArgAlaGluValIleH 193

     971 ACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAG 1020
         ::   |||   :::|||:::|||:::::::|||   :::|||   :::
     193 isAspAlaIleGlnValAlaAsnIleAlaAsp...LeuTyrPheArgThr 208

    1021 CTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGA 1070
         ||||||||||||||   :::|||   :::|||||||||   ::::::::
     209 LeuAsnThrArgValSerValValTyrIleGluThrTrpGlyLysAsnGl 225

    1071 TCAG...ATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCT 1117
         :      ||||||   :::   :::      :::   :::   :::|||:
     225 nAlaValIleAspGlySerLysAspIleSerLysAlaIleSerAsnPheA 242

    1118 CAAAATACCGGCAGCGC......ATTAAGCAGCATGCTGATGCTGTGCAC 1161
         ::   |||   :::|||      ::::::   :::   |||:::::::::
     242 snAspTyrThrSerArgAsnLeuPheGlnIleGluArgAspThrThrGln 258

    1162 CTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTT 1211
         |||::::::      ||||||          :::::::::::::::
     259 LeuLeuThrGlyGluThrPheAlaGlyGlyGluAlaGlyMetAlaValPr 275

    1212 TGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGT.............GTGA 1249
               ||||||:::      |||:::||||||             ||||
     275 oGluThrValCysThr...ProArgAlaValGlyIleSerValAspValA 291

    1250 ATGAGTATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTG 1299
         ||   |||      |||      :::|||   ::::::::::::   :::
     291 snValTyrGlu...ProHisLeuLeuAlaGlyThrMetAlaHisMetIle 306

    1300 GCTCAAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATG 1349
         ::::::|||:::|||:::      :::   :::      :::   :::||
     307 GlyHisAsnIleGlyMetGlyHisAspAspGly......ArgGluGluCy 321

    1350 TGACTGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGT 1399
         |   |||   :::   |||   ||||||||||||   ::::::   |||:
     321 sPheCysArgAsp...TrpHisGlyCysIleMetAlaGlnSerIleValG 337

    1400 CCCATTCTCGA............AAATTTTCAAAGTGCAGCATTTTGGAG 1437
         ::::::::::            |||||||||:::||||||      :::
     337 lyGlnGluAsnValGlnProTyrLysPheSerGluCysSerLysLysAsp 353

    1438 TATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCC 1487
         |||  |||   |||:::   |||   |||   ||||||:::|||:::||
     354 TyrIleAspAlaLeuArgThrGlyHisGlyLeuCysLeuLeuAsnLysPr 370

    1488 AACAAAGCTATTTGAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTG 1537
         |:::::::::            :::||||||||      ||||||
     370 oAsnGluIleGluLeuArgArgAsnCysGlyAsnLysValValGluGluA 387

    1538 GGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGC...TATGGATTATGC 1584
           ||||||||||||||||          ||||||          |||
     387 spGluGluCysAspCysGlyThrPheGluGluCysAlaLeuAspGlnCys 403

    1585 TGTAAG......AAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGG 1628
         |||           |||:::|||::::::   |||:::|||::::::||
     404 CysAspGlyIleThrCysLysLeuLysSerGluAlaGlnCysAlaSerGl 420

    1629 GCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAAT 1678
```

```
     |   ||||||:::      :::|||   ::::::|||:::   |||   |
 420 yAlaCysCysAsp......GlnCysArgLeuArgProLysAspTyrIleC 435

1679 GCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTACTGGAGAC 1728
     ||||||||:::   ||||||||||||:::   |||||||||   |||:::
 435 ysArgAspSerAsnAsnGluCysAspLeuProGluTyrCysAspGlyGlu 451

1729 TCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATGCATGCAA 1778
        ||||||||||||   ::::::   |||:::::::|||      |||::
 452 IleGlyGlnCysProSerAspValPheLysLysAsnGlySerProCysGl 468

1779 TCAAAATCAG............GGCCGCTGCTACAATGGCGAGTGCAAGA 1816
     :   ::::::            |||   |||:::::::|||   |||   |
 468 yLeuSerLysThrGlyIleSerGlyTyrCysPheGlnGlyTyrCysProT 485

1817 CCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCT 1866
     ||   :::   ||||||:::   |||||||||      ::::::::::::
 485 hrLeuSerLeuGlnCysGluAlaIleTrpGlyTyrGlyGlySerAlaAla 501

  67 GACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAA 1916
     |||:::   |||||||||:::::::|||::::::|||:::   :::|||::
  02 AspArgGlnCysTyrGluGlnPheAsnSerLysGlySerIleAsnGlyHi 518

1917 CTGCGGGAAGGAT...GGAGACCGGTGGATTCAGTGCAGCAAACATGATG 1963
     :||||||:::|||   ::::::::::::|||:::|||:::   ::::::|
 518 sCysGlyArgAspAlaAsnGluHisTyrIleLysCysGluProGluAsnV 535

1964 TGTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCA...... 2007
     ||   ||||||   |||   |||   :::      |||   |||
 535 alGlnCysGlyThrLeuGlnCysLysAspGlyGluArgGlnProValAsn 551

2008 ...CGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCA 2054
           |||   ||||||   ::::::   |||   :::
 552 AspGlyIleAspGlnLeuTyrSerArgThrIleIleSerIle........ 565

2055 TCAAGGCCGGGTGATTGACTGC......AGTGGTGCCCATGTAGTTTTAG 2098
      :::|||:::   ::::::|||      ::::::::::::|||      :
 566 .LysGlyGlnGluPheGluCysLysAlaThrSerGlyGlnValGlySerA 582

2099 ATGATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCG 2148
     :::::      :::   |||   |||:::||||||||||||||||||
 582 snSerTyrProGluHisGlyLeuValLysAspGlyThrProCysGlyAsp 598

2149 TCTATGATGTGTTTAGATCGGAAGTGCCTACAAATT...CAAGCCCTAAA 2195
     :::::::::||||||::::::   |||:::::::::         :::::
 599 AsnLeuIleCysLeuAsnGlnThrCysValSerLeuProHisValAs 615

2196 TATGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATG 2245
     ::::::::::||||||   :::|||:::|||      ||||||   ||||
  615 pGlnThrLysCysProAlaAsnSerGlnGlyLeuGluCysSerGluHisG 632

   46 GGGTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGG 2295
     ||||||||:::|||         |||:::|||||||:::   |||:::|||
   32 lyValCysThrAsnThrAsnArgCysPheCysAspMetGlyTrpGlyGly 648

2296 ACAGATTGCAGT..................................... 2307
     :::|||||||||
  649 AsnAspCysSerSerValValLeuLeuThrThrAlaLeuProThrGluAl 665

2308 .ATCCGGGATCCAGTTAGGAACCTTCACCCCCCCAAGGATGAAGGACCCA 2356
      :::      |||   :::::::::          |||   |||   |||
  665 aLeuProThrProGluAsnThrIleLysMetGluLysLysGluThrProT 682

2357 AGGGT 2361
       |||
  682 yrGly 683

seq_name: sp_invertebrate:Q9GZ15

seq_documentation_block:
ID   Q9GZ15      PRELIMINARY;      PRT;   694 AA.
AC   Q9GZ15;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MIND-MELD (FRAGMENT).
GN   MMD.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Harper C.H., Chase B.A.;
RT   "The Drosophila mind-meld gene encodes a neuronally expressed ADAM
RT   protein.";
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF252287; AAF98331.1; -.
FT   NON_TER       1      1
SQ   SEQUENCE   694 AA;  76317 MW;  37736F7B064A72E9 CRC64;


alignment_scores:
            Quality:  883.00              Length:    690
              Ratio:  2.123                 Gaps:     29
 Percent Similarity:  60.290   Percent Identity: 31.304

alignment_block:
 US-09-634-252A-2 x Q9GZ15   ..

 Align seg 1/1  to: Q9GZ15  from: 1  to: 694

  565 TACTCTAAGGGTGGA.............................GAGCA 584
      |||       ||||||                            |||||
    3 TyrHisValGlyGlyTyrLeuValAspGlyAsnArgHisAspIleGluHi 19

  585 CTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTC 634
      ||||||||||||||||:::::::::             ::::::::::|||:
   19 sCysTyrTyrHisGlyThrValLysAspTyrProGlyAlaSerAlaAlaP 36

  635 TGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATGATACCTTCGTG 684
      ::   ||||||||||||:::   |||:::::::::   :::   :::
   36 heHisThrCysAsnGlyValSerGlyValIleHisIleGlyAsnGluThr 52

  685 TATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACAGGTCG 734
      ::::::|||:::|||:::         :::            :::   ::
   53 PheValIleHisProPhe.........TyrGlyGlyAspLeuSerLysHi 66

  735 ACCACATATAATC......CAGAAAACCTTGGCAGGACAGTATTCTAAGC 778
      :||||||:::|||         :::|||   |||::::::         :
   66 sProHisValIlePheGluAlaArgThrLysAlaAsnLysGlyCysAlaA 83

  779 AAATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCT 828
      ::       ||||||                |||:::|||
   83 snSerGlyAsnLeu..............AspSerTrpArg......... 91

  829 GAATTACAGTGGTTGAAAAGAAGGAAGAGAGCAGTGAATCCATCA...CG 875
                  |||:::|||||||   :::   ::::::   :::
   92 ............LeuSerArgArgThrLysHisLeuSerAlaGlyValAl 104

  876 TGGTATATTTGAAGAAATG............................. 894
       |||:::   ||||||:::
  104 aGlyValValGluGluIleLeuGlnAlaGlyValGlyArgAsnLysArgA 121

  895 ................AAATATTTGGAACTTATGATTGTTAATGATCAC 927
                      ||||||:::|||      |||:::   |||
  121 spValArgGluAlaThrLysTyrIleGluThrAlaIleIleValAspLys 137

  928 AAAACGTAT...AAGAAGCATCGCTCTTCTCATGCACATACCAACAACTT 974
            :::   |||::::::   |||::::::|||::::::   :::
  138 AlaMetPheAspLysArgAsnGlySerThrArgAlaGluValIleHisAs 154

  975 TGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCA 1024
       |||   :::|||:::|||::::::|||   :::|||   :::   ||||
  154 pAlaIleGlnValAlaAsnIleAlaAsp...LeuTyrPheArgThrLeuA 170
```

```
1025 ACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAG 1074
     ||||||||||||   :::|||   :::|||||||||   :::::::::
 170 snThrArgValSerValValTyrIleGluThrTrpGlyLysAsnGlnAla 186

1075 ...ATTGACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAA 1121
        ||||||   :::   :::       :::   :::   :::|||:::
 187 ValIleAspGlySerLysAspIleSerLysAlaIleSerAsnPheAsnAs 203

1122 ATACCGGCAGCGC......ATTAAGCAGCATGCTGATGCTGTGCACCTCA 1165
      |||   :::|||      ::::::   :::   |||:::::::::|||:
 203 pTyrThrSerArgAsnLeuPheGlnIleGluArgAspThrThrGlnLeuL 220

1166 TCTCGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGA 1215
     :::::      ||||||          ::::::::::::::
 220 euThrGlyGluThrPheAlaGlyGlyGluAlaGlyMetAlaValProGlu 236

1216 GGTGTCTGTTCTCGCACAAGAGGAGTTGGT.....GTGAATGA 1253
        ||||||:::      |||:::||||||          ||||||
 237 ThrValCysThr...ProArgAlaValGlyIleSerValAspValAsnVa 252

1254 GTATGGTCTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTC 1303
      |||      |||      :::|||   ::::::::::::   ::::::
 252 lTyrGlu...ProHisLeuLeuAlaGlyThrMetAlaHisMetIleGlyH 268

1304 AAAACCTTGGAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATGTGAC 1353
     ::|||:::|||:::      :::   :::      :::   :::|||
 268 isAsnIleGlyMetGlyHisAspAspGly......ArgGluGluCysPhe 282

1354 TGCACAGAATCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCA 1403
     |||   :::   |||   ||||||||||   ::::::   |||:::::
 283 CysArgAsp...TrpHisGlyCysIleMetAlaGlnSerIleValGlyGl 298

1404 TTCTCGA............AAATTTTCAAAGTGCAGCATTTTGGAGTATA 1441
     :::::::           |||||||||:::||||||      :::|||
 298 nGluAsnValGlnProTyrLysPheSerGluCysSerLysLysAspTyrI 315

1442 GAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACA 1491
       |||   |||:::   |||   |||   ||||||:::|||:::|||:::
 315 leAspAlaLeuArgThrGlyHisGlyLeuCysLeuLeuAsnLysProAsn 331

1492 AAGCTATTTGAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGA 1541
     ::::::           :::|||||||||      ||||||      ||
 332 GluIleGluLeuArgArgAsnCysGlyAsnLysValValGluGluAspGl 348

1542 GGAGTGTGATTGTGGTTTTCATGTGGAATGC...TATGGATTATGCTGTA 1588
     ||||||||||||||||         ||||||            ||||||
 348 uGluCysAspCysGlyThrPheGluGluCysAlaLeuAspGlnCysCysA 365

 589 AG......AAATGTTCCCTCTCCAACGGGGCTCACTGCAGCGACGGGCCC 1632
               |||:::|||::::::   |||:::|||::::::|||
 365 spGlyIleThrCysLysLeuLysSerGluAlaGlnCysAlaSerGlyAla 381

1633 TGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGCCG 1682
     ||||||:::      :::|||   ::::::|||:::   |||   |||||
 382 CysCysAsp......GlnCysArgLeuArgProLysAspTyrIleCysAr 396

1683 GGATGCTGTGAACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTG 1732
     ||||:::   ||||||||||||:::   |||||||||   |||:::   |
 396 gAspSerAsnAsnGluCysAspLeuProGluTyrCysAspGlyGluIleG 413

1733 GTCAGTGCCCACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAA 1782
     ||||||||||   ::::::   |||::::::|||      |||:::
 413 lyGlnCysProSerAspValPheLysLysAsnGlySerProCysGlyLeu 429

1783 AATCAG............GGCCGCTGCTACAATGGCGAGTGCAAGACCAG 1820
     ::::::           |||   |||::::::|||   |||   |||
 430 SerLysThrGlyIleSerGlyTyrCysPheGlnGlyTyrCysProThrLe 446

1821 AGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTGACA 1870
      :::   ||||||:::   |||||||||      ::::::::::|||:
 446 uSerLeuGlnCysGluAlaIleTrpGlyTyrGlyGlySerAlaAlaAspA 463

1871 AGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGC 1920
     ::   |||||||||:::::|||:::::|||:::   :::|||:::|||
  463 rgGlnCysTyrGluGlnPheAsnSerLysGlySerIleAsnGlyHisCys 479

1921 GGGAAGGAT...GGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTT 1967
     |||:::|||   ::::::::::::|||:::|||:::   ::::::|||
  480 GlyArgAspAlaAsnGluHisTyrIleLysCysGluProGluAsnValGl 496

1968 CTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCA.........C 2008
      ||||||   |||   |||   :::      |||   |||
  496 nCysGlyThrLeuGlnCysLysAspGlyGluArgGlnProValAsnAspG 513

2009 GTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAA 2058
       |||   ||||||   ::::::   |||   :::              :::
  513 lyIleAspGlnLeuTyrSerArgThrIleIleSerIle.........Lys 526

2059 GGCCGGGTGATTGACTGC......AGTGGTGCCCATGTAGTTTTAGATGA 2102
     |||:::   ::::::|||      ::::::::::::|||      :::::
  527 GlyGlnGluPheGluCysLysAlaThrSerGlyGlnValGlySerAsnSe 543

2103 TGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTA 2152
     :      :::   |||   |||:::||||||||||||||||||   ::::
  543 rTyrProGluHisGlyLeuValLysAspGlyThrProCysGlyAspAsnL 560

2153 TGATGTGTTTAGATCGGAAGTGCCTACAAATT...CAAGCCCTAAATATG 2199
     :::::|||||:::::   |||:::::::::          :::::::::
  560 euIleCysLeuAsnGlnThrCysValSerLeuProHisValAspGln 576

2200 AGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGT 2249
     ::::::||||||   :::|||:::|||      ||||||   |||||||
  577 ThrLysCysProAlaAsnSerGlnGlyLeuGluCysSerGluHisGlyVa 593

2250 GTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAG 2299
     ||||:::|||          |||:::||||||:::   |||:::|||:::|
  593 lCysThrAsnThrAsnArgCysPheCysAspMetGlyTrpGlyGlyAsnA 610

2300 ATTGCAGT.....................................ATC 2310
     ||||||||                                    :::
  610 spCysSerSerValValLeuLeuThrThrAlaLeuProThrGluAlaLeu 626

2311 CGGGATCCAGTTAGGAACCTTCACCCCCCCCAAGGATGAAGGACCCAAGGG 2360
           |||   :::::::::          |||   |||   |||
  627 ProThrProGluAsnThrIleLysMetGluLysLysGluThrProTyrGl 643

2361 TCCT.........AGTGCCACCAATCTCATAATAGGC.........TCCA 2392
                    |||   |||   :::::::::|||         |||:
  643 uSerTyrHisGlySerAsnThrValPheLeuValGlyValLeuMetServ 660

2393 TCGCTGGTGCCATCCTGGTA 2412
     ::::::|||   :::::::::
  660 alValGlyPheValPheIle 666
```

seq_name: sp_vertebrate:O42595

seq_documentation_block:

```
ID   O42595      PRELIMINARY;      PRT;   873 AA.
AC   O42595;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   METALLOPROTEASE-DISINTEGRIN (METALLOPROTEASE/DISINTEGRIN XMDC9).
GN   MDC9.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Cai H., Kraetzschmar J., Alfandari D., Hunnicutt G., Blobel C.P.;
RT   "Neural crest-specific and general expression of distinct
RT   metalloprotease-disintegrins in early Xenopus laevis development.";
RL   Dev. Biol. 0:0-0(1998).
```

```
RN   [2]
RP   SEQUENCE OF 427-474 FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE=97349132; PubMed=9205136;
RA   Shilling F.M., Kratzschmar J., Cai H., Weskamp G., Gayko U.,
RA   Leibow J., Myles D.G., Nuccitelli R., Blobel C.P.;
RT   "Identification of metalloprotease/disintegrins in Xenopus laevis
RT   testis with a potential role in fertilization.";
RL   Dev. Biol. 186:155-164(1997).
CC   -!- SIMILARITY: CONTAINS A DISINTEGRIN DOMAIN.
DR   EMBL; AF032382; AAC61846.1; -.
DR   EMBL; U78187; AAB87147.1; -.
DR   HSSP; P18619; 1FVL.
DR   MEROPS; M12.209; -.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PRINTS; PR00289; DISINTEGRIN.
DR   ProDom; PD000664; -; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS00427; DISINTEGRIN_1; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR   SMART; SM00050; DISIN; 1.
KW   Integrin; Protease; Metalloprotease; Blood coagulation; Platelet;
KW   Cell adhesion; Venom.
SQ   SEQUENCE   873 AA;  95261 MW;  6EB9D72C2DBC7533 CRC64;


alignment_scores:
            Quality:  866.50              Length:    739
              Ratio:  1.892                Gaps:     31
 Percent Similarity:  61.976   Percent Identity: 31.800

alignment_block:
 US-09-634-252A-2 x O42595   ..

 Align seg 1/1  to: O42595  from: 1  to: 873

     349 CTCATATATTACATCAACCAAGACTCG...GAAAGCCCTTATCACGTTCT 395
         ||||||      |||:::|||   |||      |||  |||::::::::
      28 LeuIleGlnGlyIleGluGlnThrSerPheLeuSerSerTyrAsnIleIl 44

     396 T......GACACAAAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCC 439
         :      :::      ||||||   ::::::      :::   :::
      44 eIleProGlnLysValAlaArgAlaArgArgAspThrSerGlySerSerG 61

      40 ATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATT 489
                :::   |||:::   ||||||:::   |||:::|||   :::
      61 lyGluAsnSerLeuSerTyrAlaIleGluValGluGlyLysLysHisVal 77

     490 CTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGAT 539
         |||   |||      |||:::   :::::::::   |||::::::
      78 LeuHisLeuGluLysAsnArgAspPheIleAlaArgAspPheAlaValTy 94

     540 TCACTAC...GAAAATGGGAAA......CCACAGTACTCTAAGGGTGGAG 580
             |||   |||||||||:::         :::      :::      |
      94 rThrTyrAlaGluAsnGlySerIleIleSerSerGlnLeuGlnAspLysG 111

     581 AGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTG 630
         ||||||||:::|||:::|||      :::||||||   :::|||   |||
     111 luHisCysHisTyrGlnGlyTyrAlaGluGlyValProGluSerValVal 127

     631 GCTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATGATACCTT 680
         |||:::||||||||||:::||||||:::||||||::::::   ::::::
     128 AlaIleSerThrCysSerGlyLeuArgGlyMetLeuHisLeuGlnAsnSe 144

     681 CGTGTATATGATAGAGCCACTAGAG.....................CTGG 709
             |||   ||||||||||||:::                     ||||
     144 rSerTyrGlyIleGluProLeuAspAsnSerAspLysPheGlnHisLeuV 161

     710 TTCAT.........GATGAGAAAAGCACAGGTCGACCACATATAATCCAG 750
         ||:::         |||   ||||||      :::|||   :::
     161 alTyrArgMetGluAspValLysSer......GluProMetValCysGly 175

     751 AAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGA 800
             |||   |||   :::       :::         |||
     176 ValThrAlaAlaAspLysGluProGluThrGluLeuAsnPro........ 189

     801 AAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGAA 850
                       |||   :::::::::         |||   ||||
     190 ................ProSerMetThrGln.........LeuLeuArgA 198

     851 GGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATAT 900
         ||||||||||:::   |||                   :::   :::|||
     198 rgLysArgAlaIleLeuPro.................GlnThrArgTyr 208

     901 TTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTC 950
         :::||||||:::||||||   |||   :::   :::
     209 ValGluLeuPheIleValValAspLysGluArgPheAspMetLeuSerAr 225

     951 TTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGG 1000
          :::::::::   ::::::          ::::::   |||   :::|
     225 gAsnGluSerAlaValArgAlaGluMetValGlnLeuSerAsnTyrLeuA 242

    1001 ATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTA 1050
         |||||:::|||      :::||||||   |||:::||||||||:::::
     242 spSerMetTyr...ThrMetLeuAsnIleArgIleValLeuValGlyLeu 257

    1051 GAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCA 1100
               ||||||:::::::::::::|||   |||      :::      ::
     258 MetIleTrpThrAspSerAsnLysIleAlaIleAspGlyThrAlaGlyAs 274

    1101 GATGCTCCATGAGTTCTCAAAATACCGGCAG.....CGCATTA 1138
         :::|||   :::|||   ::::::|||:::            ::::
     274 pValLeuGlyArgPheValGlnTrpArgGluThrAsnLeuValSerLeuA 291

    1139 AGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTAT 1188
         :::::|||   |||:::::::::|||:::   ::      |||
     291 rgArgHis...AspSerAlaGlnLeuValLeuLysLysGlyPheGly... 305

    1189 AAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGG 1238
               :::::::::::::::   |||   |||||||||||:::::::::
     306 GlyThrAlaGlyMetAlaPheValGlyThrValCysSerArgSerHisAl 322

    1239 AGTTGGTGTGAAT.........GAGTATGGTCTTCCAATGGCAGTGGCAC 1279
         :   |||:::|||         :::|||::::::   ::::::   |||:
     322 aGlyGlyIleAsnValPheSerHisTyrAsnVal...GlnSerPheAlaS 338

    1280 AAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCT 1329
         ::::::::::::::::|||:::::|||||||||:::::::   :::
     338 erIleValAlaHisGluLeuGlyHisAsnLeuGlyMetAsnHisAspAsp 354

    1330 TCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCAT 1379
         :::              :::|||   |||      |||      :::|||||
     355 GlyArg.........AsnCysPheCys......SerValAspAsnCysIl 366

    1380 CATGGAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCA 1429
         ||||   ::::::|||::::::   |||:::::|||||:::||||||  -
     366 eMet...AsnSerGlyAlaThrGlySerLysAsnPheSerSerCysSerG 382

    1430 TTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTC 1479
              :::::::::   :::      :::   ||||||:::||||||:::
     382 luGluAspPheGluLysLeuThrLeuAsnLysGlyGlyThrCysLeuLeu 398

    1480 AAC......AGGCCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGG 1523
         |||      :::|||   :::   ::::::         |||||||||
     399 AsnValProLysProAspGluAlaTyrSerAlaProPheCysGlyAsnLy 415

    1524 ATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGC. 1572
```

```
      |||:::::::|||||||||||||||||||||||          ||||  |
  415 sLeuValAspValGlyGluGluCysAspCysGlyAsnProLysGluCysG 432

 1573 ..TATGGATTATGCTGTAAG......AAATGTTCCCTCTCCAACGGGGCT 1614
                 |||||||:::          |||::::||   :::||| ||
  432 luLysAspProCysCysGluProGlyThrCysLysLeuArgSerGlyAla 448

 1615 CACTGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCC 1664
      :::|||:::   |||   ||||||:::|||       |||   |||:::||
  449 GlnCysAlaTyrGlyThrCysCysGlnAsn......CysGlyPheSerPr 463

 1665 ACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAAT 1714
      |   |||       ||||||   ::::::||||||||||||:::   | ||
  463 oGlyGlyThrValCysArgAlaValAlaAsnGluCysAspLeuProGluT 480

 1715 ATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGAC 1764
      ||||||:::|||:::|||       |||   |||:::::::::   |||:::
  480 yrCysAsnGlySerSerProPheCysGlnProAspValTyrIleGlnAsn 496

 1765 GGATATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAA 1814
      |||       |||:::::::::::::::   ||||||||||||   |||::
  497 GlyAsnProCysGlnAsnSerLysAlaTyrCysTyrAsnGlyMetCysGl 513

 1815 GACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGT 1864
      ::::   |||   |||||||||   |||:::|||:::||||||   ::::
  513 nSerTyrAspAlaGlnCysGlnAlaIlePheGlySerLysAlaLysSerA 530

 1865 CTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGA 1914
      ::      :::||||||:::::::::|||:::::|||   :::   |||
  530 laProProIleCysTyrGlnGluValAsnSerLysGlyAspArgPheGly 546

 1915 AACTGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGT 1964
      |||||||||   :::|||:::   :::   :::|||:::   ::::::::
  547 AsnCysGlyPheGlnGlyAsnAspTyrArgLysCysAspThrArgAsnAl 563

 1965 GTTCTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTG 2014
      :   ||||||   |||   |||   |||:::         |||   :::|
  563 aArgCysGlyLysLeuGlnCysGluAsnValGluSerMetProValPheG 580

 2015 GTCAACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGG 2064
      ||   ::::::   :::   |||   |||   :::::::      :::
  580 ly...IleArgProSerTyrIleGlnThrProIleHisGlyThrSerThr 595

 2065 GTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACG...GA 2111
      |||       |||   |||:::          |||   :::|||:::   ||
  596 Val......CysTrpGlyValAspPheGlnLeuGlySerAspValProAs 610

  112 TGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTT 2161
      |   :::   |||::::::||||||   |||      |||   :::|||
  610 pProAlaMetValAsnGluGlyThrLysCysAspGluSerLysValCysS 627

 2162 TAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCA 2211
           :::|||::::::   :::::::|||:::      :::|||
  627 erLysPheGlnCysValAspAlaSerValLeuGlnTyr...AspCysAsp 642

 2212 CTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGA 2261
      ::::::::::|||         |||:::|||:::||||||||:::::::::
  643 ValGlnLysLys.........CysGlyGlyAsnGlyValCysAsnSerAs 656

 2262 AGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTATCC 2311
      :   :::|||   |||:::      ||||||      :::|||:::   :
  656 nLysAsnCysHisCysAsnGluGlyTrpAlaProProAsnCysGlnThrG 673

 2312 GGGATCCAGTTAGGAACCTTCACCCCCCCAAGGATGAAGGACCCAAG... 2358
      ::            ::::::           |||:::||||||
  673 luGlyTyrGlyGlySerIle...........AspSerGlyProThrTyr 685

 2359 GGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCT 2408
      :::   :::   |||:::|||      |||   ::::::      :::||
  686 AsnAspLysAspThrSerLeuArgAspGlyLeuLeuValPhePhePheLe 702

 2409 GGTAGCAGCTATTGTC...CTTGGGGGCACAGGCTGGGGATTTAAAAATG 2455
      |:::::::   ::::::   |||||||:::      :::      :::|||
```

```
  702 uIleValProLeuLeuGlyLeuGlyAlaPheValPhePheArgArgAsnG 719

 2456 TCAAGAAGAGAAGGTTC 2472
           ||||||||||||
  719 luLeuLysArgArgPhe 724

seq_name: sp_invertebrate:Q9VFH0

seq_documentation_block:
ID   Q9VFH0      PRELIMINARY;      PRT;  1569 AA.
AC   Q9VFH0;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   CG7649 PROTEIN.
GN   CG7649.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003705; AAF55088.1; -.
DR   HSSP; P18619; 1FVL.
DR   FlyBase; FBgn0038253; CG7649.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
```

```
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR   SMART; SM00050; DISIN; 1.
KW   EGF-like domain; Glycoprotein.
SQ   SEQUENCE   1569 AA;  168825 MW;  606E19E4DDEF63B2 CRC64;


alignment_scores:
            Quality:  850.00              Length:    727
              Ratio:  1.991                Gaps:     26
 Percent Similarity:  58.735   Percent Identity: 29.161

alignment_block:
 US-09-634-252A-2 x Q9VFH0    ..

  gn seg 1/1  to: Q9VFH0  from: 1  to: 1569

  379 AGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAA 428
      |||   |||   |||:::      ::::::::::|||   :::      ::
   42 SerThrTyrThrValIleArgProGlnValGluHisGlyArgThrLysAr 58

  429 TAAG.........................GCTGTCCATCTGGCCCAGG 451
      ::::                         ::::::|||      |||
   58 gSerLeuLeuThrThrLeuAspAlaThrAspGlyLeuHisThrProGlnI 75

  452 CAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTCTTGACCTCATA 501
        |||::::::              |||::::::   ::::::||||||
   75 leSerLeuSerTyrThrHisGluGlyLysArgValValValAspLeuGln 91

  502 CTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATTCACTACGAAAA 551
        |||:::   ||||||   :::::::::      :::::::|||:::||
   92 ArgAsnAspLeuLeuLeuProAsnSerHisPhe...LeuArgTyrGlnAs 107

  552 TGGGAAACCACAGTACTCTAAGGGTGGA..................... 579
      |::::::         :::   |||
  107 nAlaSerArgGlyAlaThrProGlyHisValValThrThrPheThrLysT 124

  580 ........GAGCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGAC 621
              :::   |||:::|||:::|||   |||||||||      :::
  124 hrGluValAspLeuCysHisTyrGlnGlyHisIleArgGlyLysProGlu 140

  622 TCCAAGGTGGCTCTGTCAACCTGCAATGGA...CTTCATGGCATGTTTGA 668
      |||   ||||||||||||||||||:::|||   |||   |||:::
  141 SerValValAlaLeuSerThrCysAspGlyAlaLeuAspGlyIleValIl 157

  669 AGATGATACCTTCGTGTATATGATAGAGCCACTA................ 702
       |||         :::|||:::|||:::|||
   57 eAspGlyArgGlnThrTyrPheIleHisProHisIleAspGlyArgGlyA 174

   02 ................................................. 702

  174 rgLeuGlnAspAspHisTyrLeuLeuLysGlnAlaAspMetHisProThr 190

  703 ...............................GAGCTGGTTCATGAT.. 717
                                      :::      ||||||
  191 AsnAlaThrCysGlyTyrAspAsnHisArgAspAspHisSerHisAspTy 207

  718 .GAGAAAAGCACAGGTCGA.....................CCACATATAA 745
       ||||||:::   |||                        |||   :::
  207 rGluLysAlaAspGlyAspAsnGlyLeuGlyGlyGlyIleProSerLeuP 224

  746 TCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAATGAAGAATCTCACT 795
        :::      :::|||::::::|||:::   :::
  224 roLeuArgLeuAspGlyGlyGluPheSerArgThrLeu............ 236

  796 ATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAA 845
                                                 |||::
  237 ..............................................LeuAr 238

  846 AAGAAGGAAGAGAGCAGTGAATCCATCA.........CGTGGTATATTTG 886
      ::::|||:::::::|||   :::   |||         ||||||   ::::
  238 gLysArgArgGlnAlaAspAspSerSerGlnLeuIleArgGlyProTyrA 255

  887 AAGAAATGAAA.........TATTTGGAACTTATGATTGTTAATGATCAC 927
      ::      |||         |||:::||||||:::||||||   |||:::
  255 snAlaAsnLysTyrSerSerTyrValGluLeuValIleValValAspAsn 271

  928 AAAACGTATAAGAAGCATCGCTCTTCTCATGCACATACCAACAACTTTGC 977
      |||:::||||||:::   :::::::::       :::::::::::::::
  272 LysValTyrLysAsnPheGlnGluAsnThrLysLysValHisGlnTyrCy 288

  978 AAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAGCTCAACA 1027
      :|||::::::::::|||:::::::::::::::|||         ||||||
  288 sLysGlyIleAlaAsnIleIleAsnAlaLeuTyr...ValProLeuAsnI 304

 1028 CCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATT 1077
           |||:::||||||:::|||      |||:::|||:::::::::|||
  304 lePheValAlaLeuValGlyValValIleTrpAsnGluSerAsnGluIle 320

 1078 GACATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCG 1127
      :::::::::::::::              |||::::::|||   :::|||||
  321 GluPheSerSerAspGlyAspLeuThrLeuArgAsnPheLeuAsnTyrAr 337

 1128 G......CAGCGCATTAAGCAGCATGCT...GATGCTGTGCACCTCATCT 1168
      |       :::   :::   :::|||      |||   :::::::|||::::
  337 gSerThrLysLeuValLeuAspHisProAsnAspAsnAlaGlnLeuLeuT 354

 1169 CGCGGGTGACATTTCACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGT 1218
      :::::   :::|||               :::   :::      |||
  354 hrLysGluAsnPheAlaGlyGlyValValGlyLysAlaLeuLysGlyPro 370

 1219 GTCTGTTCT......CGCACAAGAGGAGTTGGTGTGAATGAGTATGGTCT 1262
      :::|||:::        :::   ||||||:::::::::::::
  371 IleCysThrTyrGluTyrSerGlyGlyValSerMetGlnHisSerProAs 387

 1263 TCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTG 1312
       |||      ||||||   :::::::::::::::::::::::::|||:::|
  387 nProAlaMetValAlaThrThrMetAlaHisGluMetGlyHisAsnPheG 404

 1313 GAATCCAATGGGAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAA 1362
      ||:::::::   :::   |||:::             |||   |||   :::
  404 lyMetGluHisAspThrSerAsp............CysHisCysArgAsp 416

 1363 TCCTGGGGTGGCTGCATCATGGAGGAAACAGGGGTGTCCCAT...TCTCG 1409
      :::         |||:::|||      :::::::::|||
  417 Glu......LysCysValMetAlaAlaSerSerThrSerPheIleProVa 431

 1410 AAAATTTTCAAAGTGCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAG 1459
       :::::::|||:::|||||||||   :::             ::::::||||
  431 lAsnTrpSerSerCysSerIleAspGlnLeuThrIleAlaPheSerArgG 448

 1460 GAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTTGAGCCCACG 1509
      ||   :::   ||||||   |||:::|||   :::|||||||||
  448 lyMetAsnTyrCysLeuArgAsnLysProGluArgLeuPheGluSerPro 464

 1510 GAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTT 1559
        ||||||||||||:::||||||   ||||||:::||||||||||||::
  465 ThrCysGlyAsnGlyPheValGluProGlyGluGlnCysAspCysGlyLe 481

 1560 TCATGTGGAATGCTATGGATTATGCTGT......AAGAAATGTTCCCTC. 1602
      :      :::|||   :::   ||||||      :::   |||   |||
  481 uProGluHisCysGluAsnAlaCysCysAsnAlaGlnThrCysMetLeuH 498

 1603 ..TCCAACGGGGCTCACTGCAGCGACGGGCCCTGCTGTAACAATACCTCA 1650
        |||::::::|||   |||:::   |||   ||||||:::   |||:::
  498 isSerLysAsnAlaThrCysAlaThrGlyGluCysCysAspLeuThrThr 514

 1651 TGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGTG 1700
      |||      :::      |||      ||||||:::|||   ||||||||
  515 CysArgProLysLeuAlaGlySerAlaCysArgGluAlaGluAsnGluCy 531

 1701 TGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAATC 1750
```

```
      ||||:::   ||||||||||||||:::|||      ||||||   ::::
  531 sAspLeuProGluTyrCysThrGlyGluSerGluTyrCysProAlaAspV 548

 1751 TTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTAC 1800
      ::   ::::::|||         |||:::   :::|||:::   |||:::
  548 alPheArgArgAspThrGluProCysAspGlyGlyGlnAlaTyrCysPhe 564

 1801 AATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAAC 1850
      :::|||   |||:::::::::::::||||||||||:::   :::||||||
  565 HisGlyThrCysArgSerHisSerAsnGlnCysArgThrLeuTrpGlyPr 581

 1851 AAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAG 1900
         :::   :::|||:::      ||||||:::|||   ||||||||||
  581 oThrGlyAspAsnSerGlu...HisCysTyrAsnLys...AsnThrGluG 596

 1901 GCACTGAGAAGGGAAACTGCGGGAAGGAT......GGAGACCGGTGGATT 1944
      ||||||:::   ||||||||||||   :::      :::      ::::::
  596 lyThrArgLeuGlyAsnCysGlyTyrAsnArgLeuAsnLysThrPheLeu 612

 1945 CAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCT 1994
      :::|||:::::::::   |||   ||||||:::|||   |||   :::||
  613 ArgCysGluGluGlnHisValAsnCysGlyMetLeuHisCysIleHisLe 629

 1995 TACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTT 2044
      |::::::      ::::::|||:::   ::::::   ::::::      |
  629 uAsnGluArgLeuGluPheGlyMetGluSerAlaAlaValLeuSerHisS 646

 2045 CCTTCTAC...CATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTA 2091
      ||:::      |||:::   :::::::::   |||      |||   |||
  646 erTyrIleSerHisAspArgLysIleValAlaCysArgThrAlaLeuVal 662

 2092 GTTTTAGAT...GATGATACGGATGTGGGCTATGTAGAAGATGGAACGCC 2138
         |||      :::   ||||||   |||   :::   :::|||:::
  663 AspLeuGlyLeuGlnThrThrAspProGlyLeuThrProAsnGlyAlaLy 679

 2139 ATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAAG 2188
       ||||||   :::   ||||||:::|||:::::::||||||   :::::::|
  679 sCysGlyValAspLysMetCysValAspGlnArgCysLeuProValAspA 696

 2189 CCCTAAATATGAGCAGC............TGTCCACTCGATTCCAAGGGT 2226
      ||:::::::::::::::           ||||||   |||
  696 laValArgGlnLysGlyMetGlyLysProCysProGluAsp......... 709

 2227 AAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTG 2276
            |||:::|||:::|||:::|||:::::::::::::   |||   ||
  710 ......CysAsnGlyAsnGlyIleCysAsnSerArgGlyHisCysHisCy 724

  277 TGATTTCACCTGGGCAGGGACAGATTGCAGT 2307
      ||||      ::::::|||   :::||||||
  724 sAspValGlyPheGlyGlyGluSerCysSer 734
```

seq_name: sp_mammal:Q9XSL6

seq_documentation_block:

```
ID   Q9XSL6      PRELIMINARY;      PRT;   776 AA.
AC   Q9XSL6;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   EMDC II PROTEIN.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=EPIDIDYMIS;
RA   Hall L., Jury J.A.;
RT   "Sequence analysis of human and macaque eMDC II; a member of the MDC
RT   family of proteins.";
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ242014; CAB42090.1; -.
DR   HSSP; P18619; 1FVL.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
DR   Pfam; PF01562; Pep_M12B_propep; 1.
DR   PRINTS; PR00289; DISINTEGRIN.
DR   PROSITE; PS50215; ADAM_MEPRO; 1.
DR   PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR   SMART; SM00181; EGF; 1.
SQ   SEQUENCE   776 AA;  87213 MW;  08AAFE834B37F19F CRC64;
```

alignment_scores:

```
            Quality:  842.00              Length:    723
              Ratio:  1.977                 Gaps:     28
 Percent Similarity:  58.921   Percent Identity: 30.844
```

alignment_block:

US-09-634-252A-2 x Q9XSL6 ..

Align seg 1/1 to: Q9XSL6 from: 1 to: 776

```
  328 GAAATCACACTGCCTTCAAGACTCATATATTACATCAACCAAGACTCGGA 377
      |||::::::   |||   ||||||
   30 GluValValTyrProIleArgLeu.......................... 37

  378 AAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAAAAACATA 427
         |||   |||      :::::::|||      :::||||||::::::
   38 .HisProLeuHisLysArgGluValLysGluProGluGlnGlnGluGlnP 54

  428 ATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGC 477
        :::            ::::::   ::::::::::::   :::   |||
   54 heGlu...........ThrGluLeuLysTyrLysMetThrValAsnGly 66

  478 TCCAAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGA 527
      :::      :::|||   |||      |||::::::||||||:::
   67 LysIleAlaValLeuTyrLeuLysLysAsnLysAsnLeuLeuAlaProGl 83

  528 TTATGTGGAGATTCACTACGAAAAT...GGGAAA.............C 559
       |||:::|||   :::|||::::::   ||||||             |
   83 yTyrThrGluThrTyrTyrAsnSerThrGlyLysGluIleThrThrSerP 100

  560 CACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGAAGCATCAGA 609
      |||||               :::   |||||||||:::|||   |||
  100 roGlnIleMet.........AspAspCysTyrTyrGlnGlyHisIleIle 113

  610 GGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGG 659
      :::   |||||||||   :::::::::|||||||||:::||||||:::||
  114 AsnGluLysAspSerAspAlaSerIleSerThrCysArgGlyLeuArgGl 130

  660 CATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGG 709
      |   |||::::::   :::      |||:::|||||||||||||:::   :
  130 yTyrPheSerGlnGlyAsnGlnArgTyrPheIleGluProLeuSerProI 147

  710 TTCATGATGAGAAAAGCACAGGTCGACCACATATAATCCAGAAAACCTTG 759
      ::|||   :::          |||:::   |||   :::   |||
  147 leHisArgAsp.........GlyGlnGluHisAlaLeuPheLys...... 158

  760 GCAGGACAGTATTCTAAGCAAATGAAGAATCTCACTATGGAAAGAGGT.. 807
               |||:::   :::   ||||||                |||
  159 .........TyrAspProGluGluLysAsnTyrAspSerThrCysGlyTh 172

  808 .GACCAGTGGCCCTTTCTCTCTGAATTACAGTGGTTGAAAAGA...AGGA 853
       |||         ::::::   :::||||||   :::   |||
  172 rAspGlyValLeuTrpValHisAspLeuGlnAsnIleAlaArgProAlaT 189

  854 AGAGAGCAGTGAATCCATCACGTGGT...ATATTTGAAGAAATGAAATAT 900
```

```
 189 hrArgLeuValLysLeuAsnAspGlyLysValGlnLysHisGluLysTyr 205

 901 TTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTC 950
 206 IleGluTyrTyrLeuValLeuAspAsnGlyGluPheLysLysTyrAsnGl 222

 951 TTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGG 1000
 222 uAsnGlnAspGluIleArgLysArgValPheGluMetAlaAsnTyrValA 239

1001 ATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTA 1050
 239 snMetLeuTyrLys...LysLeuAsnThrHisValAlaLeuValGlyMet 254

1051 GAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCA 1100
 255 GluIleTrpThrAspGluAspLysIleAsnIleThrProAsnAlaSerPh 271

1101 GATGCTCCATGAGTTCTCAAAATACCGG.....CAGCGCATTA 1138
 271 eThrLeuGluAsnPheSerLysTrpArgGlySerValLeuProArgArgL 288

1139 AGCAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTAT 1188
 288 ysArgHis...AspIleAlaGlnLeuIleThrAlaThrGluPheAlaGly 303

1189 AAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGG 1238
 304 MetThrValGlyLeuAlaPheMetSerThrMetCysSerProTyrHisSe 320

1239 AGTTGGTGTG......AATGAGTATGGTCTTCCAATGGCAGTGGCACAAG 1282
 320 rValGlyValValGlnAspHisSerAspAsnLeuLeuArgValAlaGlyT 337

1283 TATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAACCTTCT 1332
 337 hrMetAlaHisGluMetGlyHisAsnPheGlyMetPheHisAspAsnTyr 353

1333 AGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATCAT 1382
 354 Ser.....CysLysCysProSerThrIle......CysValMe 364

1383 GGAGGAAACA...GGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCA 1429
 364 tAspLysAlaLeuSerPheTyrIleProThrAspPheSerSerCysSerA 381

1430 TTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTC 1479
 381 rgValSerTyrAspLysPhePheGluAspLysLeuSerAsnCysLeuPhe 397

  80 AACAGG......CCAACAAAGCTATTTGAGCCCACGGAATGTGGAAATGG 1523
  98 AsnAlaProLeuProThrAspIleIleSerThrProIleCysGlyAsnGl 414

1524 ATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTTCATGTGGAATGCT 1573
 414 nMetValGluMetGlyGluAspCysAspCysGlyThrSerGluGluCysT 431

1574 ATGGATTATGCTGT......AAGAAATGTTCCCTCTCCAACGGGGCTCAC 1617
 431 hrAsnIleCysCysAspAlaLysThrCysLysIleLysAlaGlyPheGln 447

1618 TGCAGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACG 1667
 448 CysThrSerGlyGluCysCysGlu......LysCysGlnPheLysLysAl 462

1668 AGGGTATGAATGCCGGGATGCTGTGAACGAGTGTGATATTACTGAATATT 1717
 462 aGlyMetValCysArgProAlaLysAspGluCysAspLeuProGluMetC 479

1718 GTACTGGAGACTCTGGTCAGTGCCCACCAAATCTTCATAAGCAAGACGGA 1767
```

```
 479 ysAspGlyLysSerGlyAsnCysProAspAspArgPheArgAlaAsnGly 495

1768 TATGCATGCAATCAAAATCAGGGCCGCTGCTACAATGGCGAGTGCAAGAC 1817
 496 PheProCysHisHisGlyLysGlyTyrCysLeuMetGlyAlaCysProTh 512

1818 CAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCAGGGTCTG 1867
 512 rLeuGlnGluGlnCysThrGluLeuTrpGlyProGlyThrLysValAlaA 529

1868 ACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAAC 1917
 529 spGlnSerCysTyrAsnArg...AsnGluGlyGlySerLysTyrGlyTyr 544

1918 TGCGGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTT 1967
 545 CysArgArgValAspAspThrLeuIleProCysLysThrAsnAspThrMe 561

1968 CTGTGGATTCTTACTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTC 2017
 561 tCysGlyLysLeuPheCys.............................G 568

2018 AACTTCAGGGTGAGATCATTCCAACTTCCTTCTACCATCAAGGCCGGGTG 2067
 568 lnGlyGlySerAspAsnLeuPro............TrpLysGlyArgIle 580

2068 ATTGACTGCAGTGGTGCCCATGTAGTTTTAGATGATGATACG......GA 2111
 581 ValThrPheLeuThrCysLysThrPheAspProGluAspThrSerGluGl 597

2112 TGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGTT 2161
 597 uIleGlyMetValAlaAsnGlyThrLysCysGlyHisAsnLysValCysI 614

2162 TAGATCGGAAGTGCCTACAAATTCAA...GCCCTAAATATGAGCAGCTGT 2208
 614 leAsnAlaGluCysValAspIleGluLysAlaTyrLysSerThrAsnCys 630

2209 CCACTCGATTCCAAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAA 2258
 631 ......SerSerLys.........CysLysGlyHisAlaValCysAspHi 642

2259 TGAAGCCACCTGCATTTGTGATTTCACCTGGGCAGGGACAGATTGCAGTA 2308
 642 sGluLeuGlnCysGlnCysGluGluGlyTrpSerProProAspCysAspA 659

2309 TCCGGGATCCAGTTAGGAACCTTCACCCCCCCAAGGATGAAGGACCCAAG 2358
 659 spSerSerValValPheTyrPhe..........................  666

2359 GGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGGTGCCATCCT 2408
 667 ......SerIleValValAlaValLeuPheProValAlaValIleSerLe 681

2409 GGTAGCAGCTATTGTCCTT 2427
 681 uValValAlaIleValIle 687

seq_name: sp_human:Q13493

seq_documentation_block:
ID   Q13493      PRELIMINARY;      PRT;   814 AA.
AC   Q13493;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   MDC15.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   TISSUE=UMBILICAL VEIN;
RX   MEDLINE=97192141; PubMed=9039960;
RA   Herren B., Raines E.W., Ross R.;
RT   "Expression of a disintegrin-like protein in cultured human vascular
RT   cells and in vivo.";
RL   FASEB J. 11:173-180(1997).
DR   EMBL; U46005; AAC51112.1; -.
DR   HSSP; P18619; 1FVL.
DR   InterPro; IPR000130; -.
DR   InterPro; IPR000561; -.
DR   InterPro; IPR001590; -.
DR   InterPro; IPR001762; -.
DR   InterPro; IPR002870; -.
DR   Pfam; PF00200; disintegrin; 1.
DR   Pfam; PF01421; Reprolysin; 1.
     Pfam; PF01562; Pep_M12B_propep; 1.
     ProDom; PD000664; -; 1.
     PROSITE; PS50215; ADAM_MEPRO; 1.
     PROSITE; PS50214; DISINTEGRIN_2; 1.
DR   PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
DR   SMART; SM00050; DISIN; 1.
SQ   SEQUENCE   814 AA;  87715 MW;  22212868BC208876 CRC64;


alignment_scores:
           Quality:  826.00            Length:    863
             Ratio:  1.796               Gaps:     32
 Percent Similarity:  53.302   Percent Identity: 26.767

alignment_block:
 US-09-634-252A-2 x Q13493   ..

 Align seg 1/1  to: Q13493  from: 1  to: 814

      91 GGCTCGGTGCCTGCCAGCGCCCCGGCCCGCACGCCGCCCTGCCGCCTGCT 140
         |||   :::   |||:::::::|||       :::   |||
      10 GlyLeuLeuGlyAlaGlySerProLeuProSerTrpPro..... 22

     141 TCTCGTCCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTCCCGGCCCCGCG 190
                         ||||||   ::::::::::::::::::   :::|
      23 ..............LeuProAsnIleGlyGlyThrGluGluGlnGlnA 34

     191 CCTGGGGGGCTGCTGCGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAA 240
         ||   :::      ||||||      |||
      34 laGluSerGluLysAlaProArgGluProLeu..... 44

     241 AAAAATTTGGGAGTCCTGGCAGATGAAGACAATACATTGCAACAGAATAG 290
                            |||   ::::::||||||::::::
      45 .....................GluProGlnValLeuGlnAspAspLe 53

     291 CAGCAGTAATATCAGTTACAGCAATGCAATGCAGAAAGAAATCACACTGC 340
                 ||||||   ::::::::::::|||   ::::::      |
      53 uPro......IleSerLeuLysLysValLeuGlnThrSerLeuProGluP 68

     341 CTTCAAGACTCATATATTACATCAACCAAGACTCGGAAAGCCCTTATCAC 390
         ||   |||:::
      68 roLeuArgIle..................................... 7

     391 GTTCTTGACACAAAGGCAAGACACCAGCAAAAACATAATAAGGCTGTCCA 440

      71 ..................................................  71

     441 TCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCCAAATTCATTC 490
                            ::::::|||      |||::::::   ||||
      72 ..................LysLeuGluLeuAspGlyAspSerHisIleL 82

     491 TTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATT 540
         ||:::|||:::   |||:::   |||:::   :::      :::   :::
      82 euGluLeuLeuGlnAsnArgGluLeuValProGlyArgProThrLeuVal 98

     541 CACTACGAA...AATGGGAAACCACAGTACTCTAAGGGTGGA......GA 581
            |||:::   :::|||            |||:::|||         ||

      99 TrpTyrGlnProAspGlyThrArgValValSerGluGlyHisThrLeuGl 115

     582 GCACTGTTACTACCATGGAAGCATCAGAGGCGTCAAAGACTCCAAGGTGG 631
         |:::|||   |||:::|||   :::||||||         |||   |||:
     115 uAsnCysCysTyrGlnGlyArgValArgGlyTyrAlaGlySerTrpValS 132

     632 CTCTGTCAACCTGCAATGGACTTCATGGCATGTTTGAAGATGATACCTTC 681
         :::::   ||||||:::|||||||:::|||
     132 erIleCysThrCysSerGlyLeuArgGly..................... 141

     682 GTGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAGAAAAGCACAGG 731
                         |||   ::::::      |||:::|||
     142 ..................LeuValValLeuThrProGluArgSer..... 150

     732 TCGACCACATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAAA 781

     150 ..................................................  150

     782 TGAAGAATCTCACTATGGAAAGAGGT.....GACCAGTGGCCC 819
                    |||:::|||:::|||            :::      |||
     151 ........TyrThrLeuGluGlnGlyProGlyAspLeuGlnGlyProPro 164

     820 TTTCTCTCTGAATTACAG.............................. 837
         ::::::|||:::::::|||
     165 IleIleSerArgIleGlnAspLeuHisLeuProGlyHisThrCysAlaLe 181

     838 ....TGGTTGAAAAGA.............................A 850
             |||   :::
     181 uSerTrpArgGluSerValHisThrGlnThrProProGluHisProLeuG 198

     851 GGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATAT 900
           :::|||   ::::::      |||   :::      |||   |||
     198 lyGlnArgHisIleArgArgArgArgAspValValThrGluThrLysThr 214

     901 TTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTC 950
         :::||||||:::||||||   ||||||:::      :::|||:::|||
     215 ValGluLeuValIleValAlaAspHisSerGluAlaGlnLysTyrArg.. 230

     951 TTCTCATGCACATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGG 1000
          :::      |||    . |||   :::   :::|||:::   |||:::|
     231 .AspPheGlnHisLeuLeuAsnArgThrLeuGluValAlaLeuLeuLeuA 247

    1001 ATTCTATTTACAAGGAGCAGCTCAACACCAGGGTTGTCCTGGTGGCTGTA 1050
         ||:::::::::::::      ||||||:::||||||:::||||||::::::
     247 spThrPhePheArg...ProLeuAsnValArgValAlaLeuValGlyLeu 262

    1051 GAGACCTGGACTGAGAAGGATCAGATTGACATCACCACCAACCCTGTGCA 1100
         |||:::||||||::::::|||   ::::::|||:::   ||||||:::
     263 GluAlaTrpThrGlnArgAspLeuValGluIleSerProAsnProAlaVa 279

    1101 GATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATTAAGCAGCATGCT. 1149
             |||::::::|||       :::|||          :::   |||
     279 lThrLeuGluAsnPheLeuHisTrpArg.........ArgAlaHisLeuL 293

    1150 .................GATGCTGTGCACCTCATCTCGCGGGTGACATTT 1182
                         |||:::::::::|||::::::   :::::::|||
     293 euProArgLeuProHisAspSerAlaGlnLeuValThrGlyThrSerPhe 309

    1183 CACTATAAGAGAAGCAGTCTGAGTTACTTTGGAGGTGTCTGTTCTCGC.. 1230
                        :::::::::      :::::::::||||||
     310 SerGlyProThrValGlyMetAlaIleGlnAsnSerIleCysSerProAs 326

    1231 ....ACAAGAGGAGTTGGTGTGAATGAGTATGGTCTTCCAATGGCAGTGG 1276
             :::   ||||||:::::::::::::            ::::::||||
     326 pPheSerGlyGlyValAsnMetAspHisSerThrSerIleLeuGlyValA 343

    1277 CACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAATGGGAA 1326
         ||:::   :::::::::::::|||::::::::::||||||::::::   :::
     343 laSerSerIleAlaHisGluLeuGlyHisSerLeuGlyLeuAspHisAsp 359

    1327 CCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTG 1376
              :::::::::      |||                      ||
     360 LeuProGlyAsnSerCysProCysProGlyProAlaProAlaLysThrCy 376
```

```
1377 CATCATGGAG...GAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGT 1423
     ||||||||||   :::|||            :::   :::||||||:::|
 376 sIleMetGluAlaSerThrAspPheLeuProGlyLeuAsnPheSerAsnC 393

1424 GCAGCATTTTGGAGTATAGAGACTTTTTACAGAGAGGAGGTGGAGCCTGC 1473
     |||||           :::      |||      |||   |||:::|||
 393 ysSerArgArgAlaLeuGluLysAlaLeuLeuAspGlyMetGlySerCys 409

1474 CTTTTCAACAGG.............CCAACAAAGCTATTTGAGCCCAC 1508
     ||||||:::|||              |||         |||
 410 LeuPheGluArgLeuProSerLeuProProMetAlaAlaPhe........ 423

1509 GGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTT 1558
         |||||||||   :::||||||   ||||||:::|||||||||||||
 424 ....CysGlyAsnMetPheValGluProGlyGluGlnCysAspCysGlyP 439

1559 TTCATGTGGAATGCTATGGATTATGCTGTAAGAAA......TGTTCCCTC 1602
     ||      :::|||         ||||||   :::      |||:::|||
 439 heLeuAspAspCysValAspProCysCysAspSerLeuThrCysGlnLeu 455

 603 TCCAACGGGGCTCACTGC...AGCGACGGGCCCTGCTGTAACAATACCTC 1649
           ||||||:::|||   ||||||||||||||||||:::|||
 456 ArgProGlyAlaGlnCysAlaSerAspGlyProCysCysGlnAsn..... 470

1650 ATGTCTTTTTCAGCCACGAGGGTATGAATGCCGGGATGCTGTGAACGAGT 1699
      |||   :::::::|||   |||::::::||||||   :::   ::::::|
 471 .CysGlnLeuArgProSerGlyTrpGlnCysArgProThrArgGlyAspC 487

1700 GTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGCCCACCAAAT 1749
     |||||:::   |||:::|||   |||||||||:::|||||||||||:::
 487 ysAspLeuProGluPheCysProGlyAspSerSerGlnCysProProAsp 503

1750 CTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTA 1799
     :::         ||||||      |||      :::|||:::   |||
 504 ValSerLeuGlyAspGlyGluProCysAlaGlyGlyGlnAlaValCysMe 520

1800 CAATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAA 1849
      :::|||:::|||   :::      :::|||||||||   :::||||||
 520 tHisGlyArgCysAlaSerTyrAlaGlnGlnCysGlnSerLeuTrpGlyP 537

1850 CAAAGGCTGCAGGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAA 1899
        |||      :::     :::|||   :::      |||||||::
 537 roGlyAlaGlnProAlaAlaProLeuCysLeuGlnThrAlaAsnThrArg 553

1900 GGCACTGAGAAGGGAAACTGCGGGAAGGAT...GGAGACCGGTGGATTCA 1946
     |||:::      |||:::||||||::::::   :::      ::::::::
 554 GlyAsnAlaPheGlySerCysGlyArgAsnProSerGlySerTyrValSe 570

1947 GTGCAGCAAACATGATGTGTTCTGTGGATTCTTACTCTGTACCAATCTTA 1996
     :|||:::   :::|||::::::||||||   |||   |||   :::
1570 rCysThrProArgAspAlaIleCysGlyGlnLeuGlnCysGlnThrGlyA 587

1997 CTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCAACTTCC 2046
           |||   :::|||:::::::::   :::::::::      :::
 587 rgThrGlnProLeuLeuGlySerIleArg...AspLeuLeuTrpGluThr 602

2047 TTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTT 2096
     :::      :::|||      ::::::||||||   :::|||      ||
 603 IleAspValAsnGlyThrGluLeuAsnCysSerTrpValHis......Le 617

2097 AGATGATGATACGGATGTGGGC.........TATGTAGAAGATGGAACGC 2137
     ||||      :::||||||:::            :::      ||||||
 617 uAspLeuGlySerAspValAlaGlnProLeuLeuThrLeuProGlyThrA 634

2138 CATGTGGCCCGTCTATGATGTGTTTAGATCGGAAGTGCCTACAAATTCAA 2187
       |||||||||:::::::::|||:::|||::::::|||   :::::::::
 634 laCysGlyProGlyLeuValCysIleAspHisArgCysGlnArgValAsp 650

2188 GCCCTAAATATGAGCAGCTGTCCACTCGATTCCAAGGGTAAAGTCTGTTC 2237
        |||:::   :::::::|||           ::::::|||   |||
 651 LeuLeuGlyAlaGlnGluCys............ArgSerLys...CysHi 662

2238 GGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGATTTCACCT 2287
      ||||||||||||||||:::::::::      |||   |||:::      |
 662 sGlyHisGlyValCysAspSerAsnArgHisCysTyrCysGluGluGlyT 679

2288 GGGCAGGGACAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCC 2337
     |||||      ||||||:::   :::
 679 rpAlaProProAspCysThrThrGln..................... 687

2338 CCCAAGGATGAAGGACCCAAGGGTCCTAGTGCC......ACCAATCTCAT 2381
                       |||:::   |||:::      |||:::|||::
 688 ..............LeuLysAlaThrSerSerLeuThrThrGlyLeuLe 699

2382 AATAGGCTCCATCGCTGGTGCCATCCTGGTAGCAGCTATTGTCCTTGGGG 2431
     :::::::   :::         :::|||:::::   :::::::||||||:
 699 uLeuSerLeuLeu.........ValLeuLeuValLeuValMetLeuGlyA 713

2432 GCACAGGCTGGGGATTTAAAAATGTCAAGAAGAGA.............. 2466
     :::::   |||         ::::::   :::|||
 713 laSerTyrTrpTyrArgAlaArgLeu***GlnArgLeuCysGlnLeuLys 729

2467 ...........AGGTTCGATCCTACTCAGCAAGGCCCC 2493
                :::::      :::|||:::||||||
 730 GlyProThrCysGlnTyrArgAlaAlaGlnSerGlyPro 742
```

Hits 1-3

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2001, 20:20:37 ; Search time 167.85 Seconds
(without alignments)
9348.383 Million cell updates/sec

Title: US-09-634-252A-2
Perfect score: 2499
Sequence: 1 atgaagccgcccggcagcag.....ctcagcaaggccccatctga 2499

[redacted]ng table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 730101 seqs, 313950809 residues

Total number of hits satisfying chosen parameters: 1460202

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 60 summaries

Database : N_Geneseq_0601:*
1: /SIDS8/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS8/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS8/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS8/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS8/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS8/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS8/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS8/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS8/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS8/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS8/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS8/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS8/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS8/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS8/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS8/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS8/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS8/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS8/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS8/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS8/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS8/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2499	100.0	2499	20	AAZ10207	cDNA encoding the
2	2499	100.0	3054	20	AAX78438	Human MDC3 cDNA.
3	1965.4	78.6	2268	21	AAZ45826	DNA encoding a dis
4	1449.2	58.0	2088	21	AAZ45827	Degenerate DNA enc
5	586.8	23.5	2604	20	AAZ10208	cDNA encoding the
6	586.8	23.5	2697	20	AAX78437	Human MDC2-beta cD
7	586.8	23.5	2805	20	AAX78436	Human MDC2-alpha c
8	575	23.0	3183	16	AAQ76122	Human fetal brain
9	530	21.2	2913	16	AAQ76121	Human fetal cerebe
10	513.2	20.5	2923	16	AAQ76120	Human fetal cerebe
11	369.2	14.8	446	21	AAC04043	Human secreted pro
12	338.4	13.5	1464	16	AAQ76119	Human fetal cerebe
13	154.8	6.2	2406	22	AAC85515	Degenerate first s
14	154.8	6.2	2439	22	AAC85516	Degenerate second
15	113.8	4.6	2848	18	AAT86011	Human alpha meltri
16	113.8	4.6	3305	20	AAZ33953	Human PRO545 nucle
17	113.8	4.6	3305	21	AAC78476	Human PRO545 (UNQ3
18	110.4	4.4	2760	18	AAT86007	Mouse alpha meltri
19	107.6	4.3	2253	18	AAT62559	Human metalloprote
20	107.6	4.3	2648	18	AAT62558	Human metalloprote
21	107.6	4.3	2757	22	AAC91177	ADAM gene #1. Hom
22	105.2	4.2	1183	18	AAT86014	Human beta meltrin
23	100.8	4.0	2820	18	AAT86008	Mouse beta meltrin
24	94.6	3.8	967	18	AAT86015	Human gamma meltri
25	91	3.6	3431	22	AAC85513	First splice varia
26	91	3.6	3468	22	AAC85514	Second splice vari
27	90.2	3.6	3186	21	AAC95499	Human secreted pro
c 28	86.6	3.5	936	22	AAF58252	Oligonucleotide D1
c 29	86.6	3.5	936	22	AAF58254	Oligonucleotide D1
c 30	86.6	3.5	936	22	AAF58257	Oligonucleotide D1
c 31	86.6	3.5	936	22	AAF58259	Oligonucleotide D2
c 32	86.6	3.5	936	22	AAF58262	Oligonucleotide D2
c 33	86.6	3.5	938	22	AAF58255	Oligonucleotide D1
34	86.6	3.5	2526	18	AAV02193	Human ADAM 12 full
35	85.8	3.4	2359	19	AAV07901	Cobra venom mocarh
36	82.8	3.3	1500	21	AAA59303	DNA encoding a sna
37	82.8	3.3	2373	21	AAA59307	DNA encoding a sna
38	81.8	3.3	2050	19	AAV07895	Cobra venom mocarh
39	81.4	3.3	936	22	AAF58252	Oligonucleotide D1
40	81.4	3.3	936	22	AAF58254	Oligonucleotide D1
41	81.4	3.3	936	22	AAF58257	Oligonucleotide D1
42	81.4	3.3	936	22	AAF58259	Oligonucleotide D2
43	81.4	3.3	936	22	AAF58262	Oligonucleotide D2
44	81.4	3.3	938	22	AAF58255	Oligonucleotide D1
45	81.2	3.2	2346	21	AAA59308	DNA encoding a sna
46	81.2	3.2	2373	21	AAA59226	cDNA encoding a hu
47	80	3.2	394	18	AAT86013	Human beta meltrin
48	79.8	3.2	1851	22	AAC85063	E. carinatus prepr
49	76.8	3.1	2580	18	AAT86009	Mouse gamma meltri
50	72.4	2.9	321	18	AAT86010	Human alpha meltri
51	71.8	2.9	1463	15	AAQ54637	Guinea pig PH-30,
52	71.6	2.9	1820	19	AAV07900	Cobra venom mocarh
53	71.6	2.9	2297	19	AAV07896	Cobra venom mocarh
54	70.6	2.8	1856	20	AAX28637	Nucleotide sequenc
55	70.6	2.8	2169	20	AAX90699	Human SVPH1-8 cDNA
56	70.6	2.8	2335	19	AAV07897	Cobra venom mocarh
57	70.4	2.8	2288	19	AAV07898	Cobra venom mocarh
58	70	2.8	2047	22	AAC91179	ADAM gene #3. Hom
59	70	2.8	2047	22	AAF33052	Human secreted pro
60	70	2.8	2309	19	AAV07899	Cobra venom mocarh

ALIGNMENTS

RESULT 1
AAZ10207
ID AAZ10207 standard; DNA; 2499 BP.
XX
AC AAZ10207;
XX
DT 29-OCT-1999 (first entry)
XX
DE cDNA encoding the human SVPH3-17 protein.
XX
KW SVPH3-17; metalloproteinase-disintegrin; human chromosome 2;
KW human chromosome 7; malignant hyperthermia susceptibility;
KW Zellweger syndrome; neonatal adrenoleukodystrophy;
KW infantile Refsum disease; progressive familial intrahepatic colchestatis;
KW mucopolysaccharidosis VII; split hand/foot malformation;
KW arrhythymogenic right ventricular displasia-4; Coppock-like cataract;
KW insulin dependent diabetes mellitus-12; lamellar type ichthyosis;
KW transient neonatal myasthenia gravis; congenital aculeiform cataract;
KW juvenile amyotrophic lateral sclerosis;

```
KW   familial paroxysmal choreoathetosis;
KW   Finnish lethal neonatal metabolic syndrome; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2499
FT                   /*tag=  a
XX
PN   WO9941388-A2.
XX
PD   19-AUG-1999;
XX
PF   11-FEB-1999;    99WO-US03016.
XX
PR   11-FEB-1998;    98US-0074310.
XX
PA   (IMMV ) IMMUNEX CORP.
XX
PI   Cerretti DP;

DR   WPI; 1999-527371/44.
DR   P-PSDB; AAY30207.

PT   DNA encoding the SVPH3-13 and SVPH3-17 proteins for detecting
PT   disease corresponding to chromosome 7, e.g. Zellweger syndrome
XX
PS   Claim 1; Page 7-8; 82pp; English.
XX
CC   The present sequence encodes a protein designated SVPH3-17, which
CC   is a member of the metalloproteinase-disintegrin family. The
CC   specification also describes SVPH3-13 proteins. Both sequences can
CC   be used to identify human chromosome 2 or 7, and to map genes on these
CC   two chromosomes, and also to identify genes associated with certain
CC   diseases, syndromes, or other human conditions associated with human
CC   chromosome 2 or 7. The disease that correspond to chromosome 7 include
CC   malignant hyperthermia susceptibility, Zellweger syndrome, neonatal
CC   adrenoleukodystrophy, infantile Refsum disease, progressive familial
CC   intrahepatic colchestatis, mucopolysaccharidosis VII, and split
CC   hand/foot malformation. Diseases associated with chromosome 2 include
CC   arrhythymogenic right ventricular displasia-4, insulin dependent diabetes
CC   mellitus-12, transient neonatal myasthenia gravis, juvenile amyotrophic
CC   lateral sclerosis, congenital aculeiform cataract, Coppock-like cataract,
CC   lamellar type ichthyosis, familial paroxysmal choreoathetosis, and
CC   Finnish lethal neonatal metabolic syndrome.
XX
SQ   Sequence 2499 BP; 677 A; 593 C; 645 G; 584 T; 0 other;


  Query Match             100.0%;  Score 2499;  DB 20;  Length 2499;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2499;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atgaagccgcccggcagcagctcgcggcagccgcccctggcgggctgcagccttgccggc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 atgaagccgcccggcagcagctcgcggcagccgcccctggcgggctgcagccttgccggc 60

Qy      61 gcttcctgcggcccccaacgcggccccgccggctcggtgcctgccagcgccccggcccgc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 gcttcctgcggcccccaacgcggccccgccggctcggtgcctgccagcgccccggcccgc 120

Qy     121 acgccgccctgccgcctgcttctcgtccttctcctgctgcctccgctcgccgcctcgtcc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 acgccgccctgccgcctgcttctcgtccttctcctgctgcctccgctcgccgcctcgtcc 180

Qy     181 cggccccgcgcctgggggctgctgcgcccagcgctccgcattggaatgaaactgcagaa 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 cggccccgcgcctgggggctgctgcgcccagcgctccgcattggaatgaaactgcagaa 240

Qy     241 aaaaatttgggagtcctggcagatgaagacaatacattgcaacagaatagcagcagtaat 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 aaaaatttgggagtcctggcagatgaagacaatacattgcaacagaatagcagcagtaat 300

Qy     301 atcagttacagcaatgcaatgcagaaagaaatcacactgccttcaagactcatatattac 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 atcagttacagcaatgcaatgcagaaagaaatcacactgccttcaagactcatatattac 360

Qy     361 atcaaccaagactcggaaagcccttatcacgttcttgacacaaaggcaagacaccagcaa 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 atcaaccaagactcggaaagcccttatcacgttcttgacacaaaggcaagacaccagcaa 420

Qy     421 aaacataataaggctgtccatctggcccaggcaagcttccagattgaagccttcggctcc 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 aaacataataaggctgtccatctggcccaggcaagcttccagattgaagccttcggctcc 480

Qy     481 aaattcattcttgacctcatactgaacaatggtttgttgtcttctgattatgtggagatt 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 aaattcattcttgacctcatactgaacaatggtttgttgtcttctgattatgtggagatt 540

Qy     541 cactacgaaaatgggaaaccacagtactctaagggtggagagcactgttactaccatgga 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 cactacgaaaatgggaaaccacagtactctaagggtggagagcactgttactaccatgga 600

Qy     601 agcatcagaggcgtcaaagactccaaggtggctctgtcaacctgcaatggacttcatggc 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 agcatcagaggcgtcaaagactccaaggtggctctgtcaacctgcaatggacttcatggc 660

Qy     661 atgtttgaagatgataccttcgtgtatatgatagagccactagagctggttcatgatgag 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 atgtttgaagatgataccttcgtgtatatgatagagccactagagctggttcatgatgag 720

Qy     721 aaaagcacaggtcgaccacatataatccagaaaaccttggcaggacagtattctaagcaa 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 aaaagcacaggtcgaccacatataatccagaaaaccttggcaggacagtattctaagcaa 780

Qy     781 atgaagaatctcactatggaaagaggtgaccagtggccctttctctctgaattacagtgg 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 atgaagaatctcactatggaaagaggtgaccagtggccctttctctctgaattacagtgg 840

Qy     841 ttgaaaagaaggaagagagcagtgaatccatcacgtggtatatttgaagaaatgaaatat 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ttgaaaagaaggaagagagcagtgaatccatcacgtggtatatttgaagaaatgaaatat 900

Qy     901 ttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgca 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 ttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgca 960

Qy     961 cataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcag 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 cataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcag 1020

Qy    1021 ctcaacaccagggttgtcctggtggctgtagagacctggactgagaaggatcagattgac 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 ctcaacaccagggttgtcctggtggctgtagagacctggactgagaaggatcagattgac 1080

Qy    1081 atcaccaccaaccctgtgcagatgctccatgagttctcaaaataccggcagcgcattaag 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 atcaccaccaaccctgtgcagatgctccatgagttctcaaaataccggcagcgcattaag 1140

Qy    1141 cagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagcagt 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 cagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagcagt 1200

Qy    1201 ctgagttactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtatggt 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 ctgagttactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtatggt 1260

Qy    1261 cttccaatggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatccaa 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 cttccaatggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatccaa 1320

Qy    1321 tgggaaccttctagcagaaagccaaaatgtgactgcacagaatcctggggtggctgcatc 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 tgggaaccttctagcagaaagccaaaatgtgactgcacagaatcctggggtggctgcatc 1380

Qy    1381 atggaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttggagtat 1440
```

```
Db   1381 atggaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttggagtat 1440

Qy   1441 agagacttttacagagaggaggtggagcctgccttttcaacaggccaacaaagctattt 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 agagacttttacagagaggaggtggagcctgccttttcaacaggccaacaaagctattt 1500

Qy   1501 gagcccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgtggtttt 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 gagcccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgtggtttt 1560

Qy   1561 catgtggaatgctatggattatgctgtaagaaatgttccctctccaacgggctcactgc 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 catgtggaatgctatggattatgctgtaagaaatgttccctctccaacgggctcactgc 1620

Qy   1621 agcgacgggccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgc 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 agcgacgggccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgc 1680

Qy   1681 cgggatgctgtgaacgagtgtgatattactgaatattgtactggagactcaggtcagtgc 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 cgggatgctgtgaacgagtgtgatattactgaatattgtactggagactcaggtcagtgc 1740

Qy   1741 ccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctac 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 ccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctac 1800

Qy   1801 aatggcgagtgcaagaccagagacaaccagtgtcagtacatctgggaaacaaaggctgca 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db   1801 aatggcgagtgcaagaccagagacaaccagtgtcagtacatctgggaaacaaaggctgca 1860

Qy   1861 gggtctgacaagttctgctatgaaaagctgaatacagaaggcactgaaaagggaaactgc 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 gggtctgacaagttctgctatgaaaagctgaatacagaaggcactgaaaagggaaactgc 1920

Qy   1921 gggaaggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtggattctta 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 gggaaggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtggattctta 1980

Qy   1981 ctctgtaccaatcttactcgagctccacgtattggtcaacttcagggcgagatcattcca 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 ctctgtaccaatcttactcgagctccacgtattggtcaacttcagggcgagatcattcca 2040

Qy   2041 acttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagttttagat 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     2041 acttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagttttagat 2100

     2101 gatgatacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatgatgtgt 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 gatgatacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatgatgtgt 2160

Qy   2161 ttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactcgattcc 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 ttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactcgattcc 2220

Qy   2221 aagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatttgtgat 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 aagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatttgtgat 2280

Qy   2281 ttcacctgggcagggacagattgcagtatccgggatccagttaggaacacttcacccccc 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 ttcacctgggcagggacagattgcagtatccgggatccagttaggaacacttcacccccc 2340

Qy   2341 aaggatgaaggacccaagggtcctagtgccaccaatctcataataggctccatcgctggt 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 aaggatgaaggacccaagggtcctagtgccaccaatctcataataggctccatcgctggt 2400

Qy   2401 gccatcctggtagcagctattgtccttgggggcacaggctggggatttaaaaatgtcaag 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 gccatcctggtagcagctattgtccttgggggcacaggctggggatttaaaaatgtcaag 2460

Qy   2461 aagagaaggttcgatcctactcagcaaggccccatctga 2499
          |||||||||||||||||||||||||||||||||||||
Db   2461 aagagaaggttcgatcctactcagcaaggccccatctga 2499
```

RESULT 2
AAX78438
ID AAX78438 standard; cDNA to mRNA; 3054 BP.
XX
AC AAX78438;
XX
DT 26-AUG-1999 (first entry)
XX
DE Human MDC3 cDNA.
XX
KW Metalloproteinase-like-disintegrin-like cysteine rich protein; human;
KW MDC2-alpha; MDC2-beta; MDC3; medical; treatment; diagnosis; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 224..2722
FT /*tag= a
FT /product= "MDC3"
XX
PN JP11155574-A.
XX
PD 15-JUN-1999.
XX
PF 01-DEC-1997; 97JP-0330020.
XX
PR 01-DEC-1997; 97JP-0330020.
XX
PA (EISA) EISAI CO LTD.
XX
DR WPI; 1999-398071/34.
DR P-PSDB; AAY25120.
XX
PT New protein belonging to MDC gene family - useful in medical
PT treatment and diagnosis
XX
PS Claim 17; Page 13-16; 17pp; Japanese.
XX
CC This invention describes novel human MDC2-alpha, MDC2-beta and MDC3
CC proteins and their encoding nucleic acids. The products of the invention
CC are useful in medical treatment and diagnosis.
XX
SQ Sequence 3054 BP; 815 A; 754 C; 803 G; 682 T; 0 other;

Query Match 100.0%; Score 2499; DB 20; Length 3054;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2499; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db    2461 aagagaaggttcgatcctactcagcaaggccccatctga 2499

Qy       1 atgaagccgcccggcagcagctcgcggcagccgcccctggcgggctgcagccttgccggc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     224 atgaagccgcccggcagcagctcgcggcagccgcccctggcgggctgcagccttgccggc 283

Qy      61 gcttcctgcggccccccaacgcggccccgccggctcggtgcctgccagcgccccggcccgc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     284 gcttcctgcggccccccaacgcggccccgccggctcggtgcctgccagcgccccggcccgc 343

Qy     121 acgccgccctgccgcctgcttctcgtccttctcctgctgcctccgctcgccgcctcgtcc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     344 acgccgccctgccgcctgcttctcgtccttctcctgctgcctccgctcgccgcctcgtcc 403

Qy     181 cggccccgcgcctgggggctgctgcgcccagcgctccgcattggaatgaaactgcagaa 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     404 cggccccgcgcctgggggctgctgcgcccagcgctccgcattggaatgaaactgcagaa 463

Qy     241 aaaaatttgggagtcctggcagatgaagacaatacattgcaacagaatagcagcagtaat 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     464 aaaaatttgggagtcctggcagatgaagacaatacattgcaacagaatagcagcagtaat 523

Qy     301 atcagttacagcaatgcaatgcagaaagaaatcacactgccttcaagactcatatattac 360
```

```
Db    524 atcagttacagcaatgcaatgcagaaagaaatcacactgccttcaagactcatatattac 583
Qy    361 atcaaccaagactcggaaagcccttatcacgttcttgacacaaaggcaagacaccagcaa 420
Db    584 atcaaccaagactcggaaagcccttatcacgttcttgacacaaaggcaagacaccagcaa 643
Qy    421 aaacataataaggctgtccatctggcccaggcaagcttccagattgaagccttcggctcc 480
Db    644 aaacataataaggctgtccatctggcccaggcaagcttccagattgaagccttcggctcc 703
Qy    481 aaattcattcttgacctcatactgaacaatggtttgttgtcttctgattatgtggagatt 540
Db    704 aaattcattcttgacctcatactgaacaatggtttgttgtcttctgattatgtggagatt 763
Qy    541 cactacgaaaatgggaaaccacagtactctaagggtggagagcactgttactaccatgga 600
Db    764 cactacgaaaatgggaaaccacagtactctaagggtggagagcactgttactaccatgga 823
Qy    601 agcatcagaggcgtcaaagactccaaggtggctctgtcaacctgcaatggacttcatggc 660
      824 agcatcagaggcgtcaaagactccaaggtggctctgtcaacctgcaatggacttcatggc 883
      661 atgtttgaagatgataccttcgtgtatatgatagagccactagagctggttcatgatgag 720
Db    884 atgtttgaagatgataccttcgtgtatatgatagagccactagagctggttcatgatgag 943
Qy    721 aaaagcacaggtcgaccacatataatccagaaaaccttggcaggacagtattctaagcaa 780
Db    944 aaaagcacaggtcgaccacatataatccagaaaaccttggcaggacagtattctaagcaa 1003
Qy    781 atgaagaatctcactatggaaagaggtgaccagtggccctttctctctgaattacagtgg 840
Db   1004 atgaagaatctcactatggaaagaggtgaccagtggccctttctctctgaattacagtgg 1063
Qy    841 ttgaaaagaaggaagagagcagtgaatccatcacgtggtatatttgaagaaatgaaatat 900
Db   1064 ttgaaaagaaggaagagagcagtgaatccatcacgtggtatatttgaagaaatgaaatat 1123
Qy    901 ttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgca 960
Db   1124 ttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgca 1183
Qy    961 cataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcag 1020
Db   1184 cataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcag 1243
Qy   1021 ctcaacaccagggttgtcctggtggctgtagagacctggactgagaaggatcagattgac 1080
Db   1244 ctcaacaccagggttgtcctggtggctgtagagacctggactgagaaggatcagattgac 1303
Qy   1081 atcaccaccaaccctgtgcagatgctccatgagttctcaaaataccggcagcgcattaag 1140
Db   1304 atcaccaccaaccctgtgcagatgctccatgagttctcaaaataccggcagcgcattaag 1363
Qy   1141 cagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagcagt 1200
Db   1364 cagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagcagt 1423
Qy   1201 ctgagttactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtatggt 1260
Db   1424 ctgagttactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtatggt 1483
Qy   1261 cttccaatggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatccaa 1320
Db   1484 cttccaatggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatccaa 1543
Qy   1321 tgggaaccttctagcagaaagccaaaatgtgactgcacagaatcctggggtggctgcatc 1380
Db   1544 tgggaaccttctagcagaaagccaaaatgtgactgcacagaatcctggggtggctgcatc 1603
Qy   1381 atggaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttggagtat 1440
```

```
Db   1604 atggaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttggagtat 1663
Qy   1441 agagacttttacagagaggaggtggagcctgccttttcaacaggccaacaaagctattt 1500
Db   1664 agagacttttacagagaggaggtggagcctgccttttcaacaggccaacaaagctattt 1723
Qy   1501 gagcccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgtggtttt 1560
Db   1724 gagcccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgtggtttt 1783
Qy   1561 catgtggaatgctatggattatgctgtaagaaatgttccctctccaacggggctcactgc 1620
Db   1784 catgtggaatgctatggattatgctgtaagaaatgttccctctccaacggggctcactgc 1843
Qy   1621 agcgacgggccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgc 1680
Db   1844 agcgacgggccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgc 1903
Qy   1681 cgggatgctgtgaacgagtgtgatattactgaatattgtactggagactctggtcagtgc 1740
Db   1904 cgggatgctgtgaacgagtgtgatattactgaatattgtactggagactctggtcagtgc 1963
Qy   1741 ccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctac 1800
Db   1964 ccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctac 2023
Qy   1801 aatggcgagtgcaagaccagagacaaccagtgtcagtacatctggggaacaaaggctgca 1860
Db   2024 aatggcgagtgcaagaccagagacaaccagtgtcagtacatctggggaacaaaggctgca 2083
Qy   1861 gggtctgacaagttctgctatgaaaagctgaatacagaaggcactgagaagggaaactgc 1920
Db   2084 gggtctgacaagttctgctatgaaaagctgaatacagaaggcactgagaagggaaactgc 2143
Qy   1921 gggaaggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtggattctta 1980
Db   2144 gggaaggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtggattctta 2203
Qy   1981 ctctgtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatcattcca 2040
Db   2204 ctctgtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatcattcca 2263
Qy   2041 acttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagttttagat 2100
Db   2264 acttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagttttagat 2323
Qy   2101 gatgatacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatgatgtgt 2160
Db   2324 gatgatacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatgatgtgt 2383
Qy   2161 ttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactcgattcc 2220
Db   2384 ttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactcgattcc 2443
Qy   2221 aagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatttgtgat 2280
Db   2444 aagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatttgtgat 2503
Qy   2281 ttcacctgggcagggacagattgcagtatccgggatccagttaggaaccttcacccccc 2340
Db   2504 ttcacctgggcagggacagattgcagtatccgggatccagttaggaaccttcacccccc 2563
Qy   2341 aaggatgaaggacccaagggtcctagtgccaccaatctcataataggctccatcgctggt 2400
Db   2564 aaggatgaaggacccaagggtcctagtgccaccaatctcataataggctccatcgctggt 2623
Qy   2401 gccatcctggtagcagctattgtccttgggggcacaggctgggggatttaaaaatgtcaag 2460
Db   2624 gccatcctggtagcagctattgtccttgggggcacaggctgggggatttaaaaatgtcaag 2683
Qy   2461 aagagaaggttcgatcctactcagcaaggccccatctga 2499
Db   2684 aagagaaggttcgatcctactcagcaaggccccatctga 2722
```

RESULT 3
AAZ45826
ID AAZ45826 standard; DNA; 2268 BP.
XX
AC AAZ45826;
XX
DT 25-APR-2000 (first entry)
XX
DE DNA encoding a disintegrin homologue designated zdintl.
XX
KW Human; disintegrin homologue; zdintl; cardiac myocyte; adipocyte;
KW gene therapy; cell-cell interaction; chondro sacroma; atherosclerosis;
KW Alzheimer's disease; restenosis; ischemic reperfusion; obesity;
KW intimal hyperplasia; tumour; platelet aggregation; apoptosis;
KW neurogenesis; myogenesis; arthritis; myogenesis; neurogenesis;
KW connective tissue disorder; chondrogenesis; tumour proliferation;
KW inflammation; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 3..2093
FT /*tag= a
FT /product= "zdintl"
XX
PN WO200002912-A2
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US15638.
XX
PR 10-JUL-1998; 98US-0113883.
XX
PA (ZYMO) ZYMOGENETICS INC.
XX
PI Sheppard PO, Baindur N, Deisher TA, Bishop PD;
XX
DR WPI; 2000-160898/14.
DR P-PSDB; AAY54457.
XX
PT Polypeptide useful in modulating cell-cell interaction in tissues of
PT heart, brain, spinal cord and treating chondro sarcoma,
PT atherosclerosis, restenosis, obesity, intimal hyperplasia and tumors -

Example 1; Page 123-127; 132pp; English.

The present sequence encodes a human disintegrin homologue, designated
CC zdintl. The zdintl polypeptide is a cardiac myocyte proliferation and
CC differentiation stimulator, as well as an adipocyte proliferation and
CC differentiation inhibitor. Polynucleotides encoding zdintl are used in
CC gene therapy. The zdintl polypeptide is useful in modulating cell-cell
CC interactions of cells derived from tissues of heart, brain, spinal cord
CC and skeletal muscle. It is useful in treating and diagnosing chondro
CC sacroma, atherosclerosis, Alzheimer's disease, restenosis, ischemic
CC reperfusion, obesity, intimal hyperplasia and tumors of heart, brain,
CC and spinal cord. The zdintl polypeptide is also useful in identifying
CC its new family members, antagonists, agonists and antibodies.
CC Antagonists, antibodies and fusion proteins of zdintl are useful in
CC inhibiting platelet aggregation, apoptosis, neurogenesis and myogenesis.
CC Agonists and antagonists are useful in studying cell-cell interactions,
CC arthritis, myogenesis, neurogenesis, connective tissue disorders,
CC chondrogenesis, tumour proliferation and suppression, extracellular
CC matrix proteins, repair and remodelling of ischemic reperfusion,
CC inflammation, and apoptosis.
XX
SQ Sequence 2268 BP; 654 A; 457 C; 558 G; 598 T; 1 other;

Query Match 78.6%; Score 1965.4; DB 21; Length 2268;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1969; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy    394 cttgacacaaaggcaagacaccagcaaaaacataataaggctgtccatctggcccaggca 453
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     27 cttgacacaaaggcaagacaccagcaaaaacataataaggctgtccatctggcccaggca 86

Qy    454 agcttccagattgaagccttcggctccaaattcattcttgacctcatactgaacaatggt 513
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     87 agcttccagattgaagccttcggctccaaattcattcttgacctcatactgaacaatggt 146

Qy    514 ttgttgtcttctgattatgtggagattcactacgaaaatgggaaaccacagtactctaag 573
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    147 ttgttgtcttctgattatgtggagattcactacgaaaatgggaaaccacagtactctaag 206

Qy    574 ggtggagagcactgttactaccatggaagcatcagaggcgtcaaagactccaaggtggct 633
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    207 ggtggagagcactgttactaccatggaagcatcagaggcgtcaaagactccaaggtggct 266

Qy    634 ctgtcaacctgcaatggacttcatggcatgtttgaagatgataccttcgtgtatatgata 693
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    267 ctgtcaacctgcaatggacttcatggcatgtttgaagatgataccttcgtgtatatgata 326

Qy    694 gagccactagagctggttcatgatgagaaaagcacaggtcgaccacatataatccagaaa 753
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    327 gagccactagagctggttcatgatgagaaaagcacaggtcgaccacatataatccagaaa 386

Qy    754 accttggcaggacagtattctaagcaaatgaagaatctcactatggaaagaggtgaccag 813
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    387 accttggcaggacagtattctaagcaaatgaagaatctcactatggaaagaggtgaccag 446

Qy    814 tggccctttctctctgaattacagtggttgaaaagaaggaagagagcagtgaatccatca 873
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    447 tggccctttctctctgaattacagtggttgaaaagaaggaagagagcagtgaatccatca 506

Qy    874 cgtggtatatttgaagaaatgaaatatttggaacttatgattgttaatgatcacaaaacg 933
          |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db    507 cgtggtatatttgaagaaatgaaatatttggaacttatgattggtaatgatcacaaaacg 566

Qy    934 tataagaagcatcgctcttctcatgcacataccaacaactttgcaaagtccgtggtcaac 993
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    567 tataagaagcatcgctcttctcatgcacataccaacaactttgcaaagtccgtggtcaac 626

Qy    994 cttgtggattctatttacaaggagcagctcaacaccagggttgtcctggtggctgtagag 1053
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    627 cttgtggattctatttacaaggagcagctcaacaccagggttgtcctggtggctgtagag 686

Qy   1054 acctggactgagaaggatcagattgacatcaccaccaaccctgtgcagatgctccatgag 1113
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    687 acctggactgagaaggatcagattgacatcaccaccaaccctgtgcagatgctccatgag 746

Qy   1114 ttctcaaaataccggcagcgcattaagcagcatgctgatgctgtgcacctcatctcgcgg 1173
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    747 ttctcaaaataccggcagcgcattaagcagcatgctgatgctgtgcacctcatctcgcgg 806

Qy   1174 gtgacatttcactataagagaagcagtctgagttactttggaggtgtctgttctcgcaca 1233
          |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    807 gtgacatttcactataagagaagcagtctgagttactttgaaggtgtctgttctcgcaca 866

Qy   1234 agaggagttggtgtgaatgagtatggtcttccaatggcagtggcacaagtattatcgcag 1293
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    867 agaggagttggtgtgaatgagtatggtcttccaatggcagtggcacaagtattatcgcag 926

Qy   1294 agcctggctcaaaaccttggaatccaatgggaaccttctagcagaaagccaaaatgtgac 1353
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    927 agcctggctcaaaaccttggaatccaatgggaaccttctagcagaaagccaaaatgtgac 986

Qy   1354 tgcacagaatcctggggtggctgcatcatggaggaaacaggggtgtcccattctcgaaaa 1413
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    987 tgcacagaatcctggggtggctgcatcatggaggaaacaggggtgtcccattctcgaaaa 1046

Qy   1414 ttttcaaagtgcagcattttggagtatagagacttttttacagagaggaggtggagcctgc 1473
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1047 ttttcaaagtgcagcattttggagtatagagacttttttacagagaggaggtggagcctgc 1106
```

```
Qy   1474 cttttcaacaggccaacaaagctatttgagcccacggaatgtggaaatggatacgtggaa 1533
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1107 cttttcaacaggccaacaaagctatttgagcccacggaatgtggaaatggatacgtggaa 1166

Qy   1534 gctggggaggagtgtgattgtggttttcatgtggaatgctatggattatgctgtaagaaa 1593
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1167 gctggggaggagtgtgattgtggttttcatgtggaatgctatggattatgctgtaagaaa 1226

Qy   1594 tgttccctctccaacggggctcactgcagcgacgggccctgctgtaacaatacctcatgt 1653
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1227 tgttccctctccaacggggctcactgcagcgacgggccctgctgtaacaatacctcatgt 1286

Qy   1654 ctttttcagccacgagggtatgaatgccgggatgctgtgaacgagtgtgatattactgaa 1713
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1287 ctttttcagccacgagggtatgaatgccgggatgctgtgaacgagtgtgatattactgaa 1346

Qy   1714 tattgtactggagactctggtcagtgcccaccaaatcttcataagcaagacggatatgca 1773
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1347 tattgtactggagactctggtcagtgcccaccaaatcttcataagcaagacggatatgca 1406

Qy   1774 tgcaatcaaaatcagggccgctgctacaatggcgagtgcaagaccagagacaaccagtgt 1833
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1407 tgcaatcaaaatcagggccgctgctacaatggcgagtgcaagaccagagacaaccagtgt 1466

Qy   1834 cagtacatctggggaacaaaggctgcagggtctgacaagttctgctatgaaaagctgaat 1893
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1467 cagtacatctggggaacaaaggctgcagggtctgacaagttctgctatgaaaagctgaat 1526

Qy   1894 acagaaggcactgagaagggaaactgcgggaaggatggagaccggtggattcagtgcagc 1953
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1527 acagaaggcactgagaagggaaactgcgggaaggatggagaccggtggattcagtgcagc 1586

Qy   1954 aaacatgatgtgttctgtggattcttactctgtaccaatcttactcgagctccacgtatt 2013
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1587 aaacatgatgtgttctgtggattcttactctgtaccaatcttactcgagctccacgtatt 1646

Qy   2014 ggtcaacttcagggtgagatcattccaacttccttctaccatcaaggccgggtgattgac 2073
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1647 ggtcaacttcagggtgagatcattccaacttccttctaccatcaaggccgggtgattgac 1706

Qy   2074 tgcagtggtgcccatgtagttttagatgatgatacggatgtgggctatgtagaagatgga 2133
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1707 tgcagtggtgcccatgtagttttagatgatgatacggatgtgggctatgtagaagatgga 1766

Qy   2134 acgccatgtggcccgtctatgatgtgtttagatcggaagtgcctacaaattcaagcccta 2193
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1767 acgccatgtggcccgtctatgatgtgtttagatcggaagtgcctacaaattcaagcccta 1826

Qy   2194 aatatgagcagctgtccactcgattccaagggtaaagtctgttcgggccatggggtgtgt 2253
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1827 aatatgagcagctgtccactcgattccaagggtaaagtctgttcgggccatggggtgtgt 1886

Qy   2254 agtaatgaagccacctgcatttgtgatttcacctgggcagggacagattgcagtatccgg 2313
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1887 agtaatgaagccacctgcatttgtgatttcacctgggcagggacagattgcagtatccgg 1946

Qy   2314 gatccagttaggaaccttcacccccccaaggatgaaggacccaagggtcctagtg 2368
          ||||||||||||||||||||||||||||||||||||||||||||||||||||    |||
Db   1947 gatccagttaggaaccttcacccccccaaggatgaaggacccaagggtttgtgtg 2001
```

RESULT 4
AAZ45827
ID AAZ45827 standard; DNA; 2088 BP.
XX
AC AAZ45827;
XX
DT 25-APR-2000 (first entry)
XX
DE Degenerate DNA encoding a disintegrin homologue designated zdintl.
XX
KW Human; disintegrin homologue; zdintl; cardiac myocyte; adipocyte;
KW gene therapy; cell-cell interaction; chondro sacroma; atherosclerosis;
KW Alzheimer's disease; restenosis; ischemic reperfusion; obesity;
KW intimal hyperplasia; tumour; platelet aggregation; apoptosis;
KW neurogenesis; myogenesis; arthritis; myogenesis; neurogenesis;
KW connective tissue disorder; chondrogenesis; tumour proliferation;
KW inflammation; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO200002912-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US15638.
XX
PR 10-JUL-1998; 98US-0113883.
XX
PA (ZYMO) ZYMOGENETICS INC.
XX
PI Sheppard PO, Baindur N, Deisher TA, Bishop PD;
XX
DR WPI; 2000-160898/14.
XX
PT Polypeptide useful in modulating cell-cell interaction in tissues of
PT heart, brain, spinal cord and treating chondro sarcoma,
PT atherosclerosis, restenosis, obesity, intimal hyperplasia and tumors -
XX
PS Disclosure; Page 129-130; 132pp; English.
XX
CC The present sequence represents a degenerate sequence which encodes
CC a human disintegrin homologue, designated zdintl. The zdintl polypeptide
CC is a cardiac myocyte proliferation and differentiation stimulator, as
CC well as an adipocyte proliferation and differentiation inhibitor.
CC Polynucleotides encoding zdintl are used in gene therapy. The zdintl
CC polypeptide is useful in modulating cell-cell interactions of cells
CC derived from tissues of heart, brain, spinal cord and skeletal muscle.
CC It is useful in treating and diagnosing chondro sacroma, atherosclerosis,
CC Alzheimer's disease, restenosis, ischemic reperfusion, obesity, intimal
CC hyperplasia and tumors of heart, brain, and spinal cord. The zdintl
CC polypeptide is also useful in identifying its new family members,
CC antagonists, agonists and antibodies. Antagonists, antibodies and fusion
CC proteins of zdintl are useful in inhibiting platelet aggregation,
CC apoptosis, neurogenesis and myogenesis. Agonists and antagonists are
CC useful in studying cell-cell interactions, arthritis, myogenesis,
CC neurogenesis, connective tissue disorders, chondrogenesis, tumour
CC proliferation and suppression, extracellular matrix proteins, repair
CC and remodelling of ischemic reperfusion, inflammation, and apoptosis.
XX
SQ Sequence 2088 BP; 412 A; 178 C; 381 G; 262 T; 855 other;

Query Match 58.0%; Score 1449.2; DB 21; Length 2088;
Best Local Similarity 58.4%; Pred. No. 0;
Matches 1161; Conservative 509; Mismatches 317; Indels 0; Gaps 0;

```
Qy    394 cttgacacaaaggcaagacaccagcaaaaacataataaggctgtccatctggcccaggca 453
          :| ||:|| ||:|| :| ||:||:||:||:||:||:|| || ||::| || ||:||
Db     25 ytngayacnaargcnmgncaycarcaraarcayaayaargcngtncayytngcncargcn 84

Qy    454 agcttccagattgaagccttcggctccaaattcattcttgacctcatactgaacaatggt 513
          :: ||:||:||:||:|| ||:|| :: ||:||:||::| ||::| ||::| ||:||:||
Db     85 wsnttycarathgargcnttyggnwsnaarttyathytngayytnathytnaayaayggn 144

Qy    514 ttgttgtcttctgattatgtggagattcactacgaaaatgggaaaccacagtactctaag 573
          :| :| :: :: ||:||:|| ||:||:||:||:||:|| ||:|| ||:||::: ||:
Db    145 ytnytnwsnwsngaytaygtngarathcaytaygaraayggnaarccncartaywsnaar 204

Qy    574 ggtggagagcactgttactaccatggaagcatcagaggcgtcaaagactccaaggtggct 633
          || || ||:||:||:||:||:||:|| :: ||::| || || ||:||::: ||:|| ||
Db    205 ggnggngarcaytgytaytaycayggnwsnathmgnggngtnaargaywsnaargtngcn 264

Qy    634 ctgtcaacctgcaatggacttcatggcatgtttgaagatgataccttcgtgtatatgata 693
          :| :: || ||:||:|| :| ||:|| |||||:||:||:||:|| ||:|| ||:||||:
```

```
Db    265 ytnwsnacntgyaayggnytncayggnatgttygargaygayacnttygtntayatgath 324

Qy    694 gagccactagagctggttcatgatgagaaaagcacaggtcgaccacatataatccagaaa 753
          ||:|| :| ||:::| || ||:||:||:||::: || || :| || ||:||:||:||:||:
Db    325 garccnytngarytngtncaygaygaraarwsnacnggnmgnccncayathathcaraar 384

Qy    754 accttggcaggacagtattctaagcaaatgaagaatctcactatggaaagaggtgaccag 813
          || :| || || ||:||::: ||:||:|||||:||::| || |||||::|.|| ||:||:
Db    385 acnytngcnggncartaywsnaarcaratgaaraayytnacnatggarmgnggngaycar 444

Qy    814 tggccctttctctctgaattacagtggttgaaaagaaggaagagagcagtgaatccatca 873
          ||||| ||::| :: ||::| ||:|||:| ||::| :| ||::| || || ||:|| ::
Db    445 tggccnttyytnwsngarytncartggytnaarmgnmgnaarmgngcngtnaayccnwsn 504

      874 cgtggtatatttgaagaaatgaaatatttggaacttatgattgttaatgaccacaaaacg 933
          :| || ||:||:||:||:|||||:||::| ||::| |||||:|  ||:||:||:||:||
      505 mgnggnathttygargaratgaartayytngarytnatgathggnaaygaycayaaracn 564

Qy    934 tataagaagcatcgctcttctcatgcacataccaacaactttgcaaagtccgtggtcaac 993
          ||:||:||:||::| :: :: ||:|| ||:|| ||:||:||:|| ||::: || || ||:
Db    565 tayaaraarcaymgnwsnwsncaygcncayacnaayaayttygcnaarwsngtngtnaay 624

Qy    994 cttgtggattctatttacaaggagcagctcaacaccagggttgtcctggtggctgtagag 1053
          :| || ||::: ||:||:||:||:||::| ||:|| :| || || :| || || || ||:
Db    625 ytngtngaywsnathtayaargarcarytnaayacnmgngtngtnytngtngcngtngar 684

Qy   1054 acctggactgagaaggatcagattgacatcaccaccaaccctgtgcagatgctccatgag 1113
          || ||||| ||:||:||:||:||:||:||:|| || ||:|| || ||:|||:| ||:||:
Db    685 acntggacngaraargaycarathgayathacnacnaayccngtncaratgytncaygar 744

Qy   1114 ttctcaaaataccggcagcgcattaagcagcatgctgatgctgtgcacctcatctcgcgg 1173
          ||::: ||:||::| ||::| ||:||:||:||:|| ||:|| || ||::| ||::: :|
Db    745 ttywsnaartaymgncarmgnathaarcarcaygcngaygcngtncayytnathwsnmgn 804

Qy   1174 gtgacatttcactataagagaagcagtctgagttactttggaggtgtctgttctcgcaca 1233
          || || ||:||:||:||:| :: :: :| :: ||:||:| :|| || ||::: :| ||
Db    805 gtnacnttycaytayaarmgnwsnwsnytnwsntayttygarggngtntgywsnmgnacn 864

Qy   1234 agaggagttggtgtgaatgagtatggtcttccaatggcagtggcacaagtattatcgcag 1293
          :| || || || || ||:||:||:|| :| || ||||| || || ||:|| :| :: ||:
Db    865 mgnggngtnggngtnaaygartayggnytnccnatggcngtngcncargtnytnwsncar 924

Qy   1294 agcctggctcaaaaccttggaatccaatgggaaccttctagcagaaagccaaaatgtgac 1353
          :: :| || ||:||::| || ||:||:|||||:|| :: :: :| || || ||:||:||:
Db    925 wsnytngcncaraayytnggnathcartgggarccnwsnwsnmgnaarccnaartgygay 984

Qy   1354 tgcacagaatcctggggtggctgcatcatggaggaaacaggggtgtccattctcgaaaa 1413
          ||:|| ||::: ||||| || ||:||:|||||:||:|| || || :: ||::: :| ||:
Db    985 tgyacngarwsntggggnggntgyathatggargaracnggngtnwsncaywsnmgnaar 1044

Qy   1414 ttttcaaagtgcagcattttggagtatagagacttttttacagagaggaggtggagcctgc 1473
          ||::: ||:||::: ||::| ||:||::| ||:||::| ||::| || || || || ||:
Db   1045 ttywsnaartgywsnathytngartaymgngayttyytncarmgnggrggnggngcntgy 1104

Qy   1474 cttttcaacaggccaacaaagctatttgagcccacggaatgtggaaatggatacgtggaa 1533
          :| ||:||::| || || ||::| ||:||:|| || ||:||:|| ||:|| ||:|| ||:
Db   1105 ytnttyaaymgnccnacnaarytnttygarccnacngartgyggnaayggntaygtngar 1164

Qy   1534 gctggggaggagtgtgattgtggttttcatgtggaatgctatggattatgctgtaagaaa 1593
          || || ||:||:||:||:|| ||:||:|| ||:||:||:|| :| ||:||:||:||:
Db   1165 gcnggngargartgygaytgyggnttycaygtngartgytayggnytntgytgyaaraar 1224

Qy   1594 tgttccctctccaacggggctcactgcagcgacgggccctgctgtaacaatacctcatgt 1653
          ||::: :| :: ||:|| || ||:||::: ||:|| || ||:||:||:||:|| :: ||:
Db   1225 tgywsnytnwsnaayggngncnaytgywsngayggnccntgytgyaayaayacnwsntgy 1284

Qy   1654 ctttttcagccacgagggtatgaatgccgggatgctgtgaacgagtgtgatattactgaa 1713
          :| ||:||:|| :| || ||:||:||::| ||:|| || ||:||:||:||:||:|| ||:
Db   1285 ytnttycarccnmgnggntaygartgymgngaygcngtnaaygartgygayathacngar 1344

Qy   1714 tattgtactggagactctggtcagtgcccaccaaatcttcataagcaagacggatatgca 1773
          ||:||:|| || ||::: || ||:||:|| || ||::| ||:||:||:||:|| ||:||
Db   1345 taytgyacnggngaywsnggncartgyccnccnaayytncayaarcargayggntaygcn 1404

Qy   1774 tgcaatcaaaatcagggccgctgctacaatggcgagtgcaagaccagagacaaccagtgt 1833
          ||:||:||:||:||:|| :| ||:||:||:|| ||:||:||:|| :| ||:||:||:||:
Db   1405 tgyaaycaraaycarggnmgntgytayaayggngartgyaaracnmgngayaaycartgy 1464

Qy   1834 cagtacatctggggaacaaaggctgcagggtctgacaagttctgctatgaaaagctgaat 1893
          ||:||:||:||||| || ||:|| || || :: ||:||:||:||:||:||:||::| ||:
Db   1465 cartayathtggggnacnaargcngcnggnwsngayaarttytgytaygaraarytnaay 1524

Qy   1894 acagaaggcactgagaagggaaactgcgggaaggatggagaccggtggattcagtgcagc 1953
          || ||:|| || ||:||:|| ||:||:|| ||:||:|| ||::| |||||:||:|||:::
Db   1525 acngarggnacngaraarggnaaytgyggnaargayggngaymgntggathcartgywsn 1584

Qy   1954 aaacatgatgtgttctgtggattcttactctgtaccaatcttactcgagctccacgtatt 2013
          ||:||:||:|| ||:||:|| ||::| :| ||:|| ||::| || :| || || :| ||:
Db   1585 aarcaygaygtnttytgyggnttyytnytntgyacnaayytnacnmgngcnccnmgnath 1644

Qy   2014 ggtcaacttcagggtgagatcattccaacttccttctaccatcaaggccgggtgattgac 2073
          || |'::| ||:|| ||:||:||:|| || :: ||:||:||:||:|| :| || ||:||:
Db   1645 ggncarytncarggngarathathccnacnwsnttytaycaycarggnmgngtnathgay 1704

Qy   2074 tgcagtggtgcccatgtagttttagatgatgatacggatgtgggctatgtagaagatgga 2133
          ||::: || || ||:|| || :| ||:||:||:|| ||:|| || ||:|| ||:||:||
Db   1705 tgywsnggngcncaygtngtnytngaygaygayacngaygtnggntaygtngargayggn 1764

Qy   2134 acgccatgtggcccgtctatgatgtgtttagatcggaagtgcctacaaattcaagcccta 2193
          || || ||:|| || :: ||||||||:::| ||:::| ||:||:::| ||:||:||:|| :|
Db   1765 acnccntgyggnccnwsnatgatgtgyytngaymgnaartgyytncarathcargcnytn 1824

Qy   2194 aatatgagcagctgtccactcgattccaagggtaaagtctgttcgggccatggggtgtgt 2253
          ||:||:: :: ||:|| :| ||::: ||:|| ||:|| ||::: || ||:|| || ||:
Db   1825 aayatgwsnwsntgyccnytngaywsnaarggnaargtntgywsnggncayggngtntgy 1884

Qy   2254 agtaatgaagccacctgcatttgtgatttcacctgggcagggacagattgcagtatccgg 2313
          :: ||:||:|| || ||:||:||:||:|| ||||| || || ||:||:|| ||:||:||:
Db   1885 wsnaaygargcnacntgyathtgygayttyacntgggcnggnacngaytgywsnathmgn 1944

Qy   2314 gatccagttaggaaccttcacccccccaaggatgaaggacccaagggtcctagtgccacc 2373
          ||:|| || :| ||::| ||:|| || ||:||:||:|| || ||:|| :   |:| :  :
Db   1945 gayccngtnmgnaayytncayccnccnaargaygargggnccnaarggnytntgygaytty 2004

Qy   2374 aatctca 2380
          |:|
Db   2005 ggnttya 2011


RESULT   5
AAZ10208
ID   AAZ10208 standard; DNA; 2604 BP.
XX
AC   AAZ10208;
XX
DT   29-OCT-1999  (first entry)
XX
DE   cDNA encoding the human SVPH3-13 protein.
XX
KW   SVPH3-13; metalloproteinase-disintegrin; human chromosome 2;
KW   human chromosome 7; malignant hyperthermia susceptibility;
KW   Zellweger syndrome; neonatal adrenoleukodystrophy;
KW   infantile Refsum disease; progressive familial intrahepatic colchestatis;
KW   mucopolysaccharidosis VII; split hand/foot malformation;
KW   arrhythymogenic right ventricular displasia-4; Coppock-like cataract;
KW   insulin dependent diabetes mellitus-12; lamellar type ichthyosis;
KW   transient neonatal myasthenia gravis; congenital aculeiform cataract;
KW   juvenile amyotrophic lateral sclerosis;
KW   familial paroxysmal choreoathetosis;
KW   Finnish lethal neonatal metabolic syndrome; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2604
FT                   /*tag=  a
```

```
FT                  /transl_except= (pos; 2227..2229, aa: Cys)
XX
PN   WO9941388-A2.
XX
PD   19-AUG-1999.
XX
PF   11-FEB-1999;   99WO-US03016.
XX
PR   11-FEB-1998;   98US-0074310.
XX
PA   (IMMV ) IMMUNEX CORP.
XX
PI   Cerretti DP;
XX
DR   WPI; 1999-527371/44.
DR   P-PSDB; AAY30208.
XX
PT   DNA encoding the SVPH3-13 and SVPH3-17 proteins for detecting
PT   disease corresponding to chromosome 7, e.g. Zellweger syndrome
XX
PS   Claim 1; Page 9-10; 82pp; English.
XX
CC   The present sequence encodes a protein designated SVPH3-13, which
CC   is a member of the metalloproteinase-disintegrin family. The
CC   specification also describes SVPH3-17 proteins. Both sequences can
CC   be used to identify human chromosome 2 or 7, and to map genes on these
CC   two chromosomes, and also to identify genes associated with certain
CC   diseases, syndromes, or other human conditions associated with human
CC   chromosome 2 or 7. The disease that correspond to chromosome 7 include
CC   malignant hyperthermia susceptibility, Zellweger syndrome, neonatal
CC   adrenoleukodystrophy, infantile Refsum disease, progressive familial
CC   intrahepatic colchestatis, mucopolysaccharidosis VII, and split
CC   hand/foot malformation. Diseases associated with chromosome 2 include
CC   arrhythymogenic right ventricular displasia-4, insulin dependent diabetes
CC   mellitus-12, transient neonatal myasthenia gravis, juvenile amyotrophic
CC   lateral sclerosis, congenital aculeiform cataract, Coppock-like cataract,
CC   lamellar type ichthyosis, familial paroxysmal choreoathetosis, and
CC   Finnish lethal neonatal metabolic syndrome.
XX
SQ   Sequence 2604 BP; 782 A; 510 C; 633 G; 679 T; 0 other;

  Query Match             23.5%;  Score 586.8;  DB 20;  Length 2604;
  Best Local Similarity   57.2%;  Pred. No. 5.5e-150;
  Matches 1237;  Conservative    0;  Mismatches  887;  Indels   38;  Gaps    8;

Qy    320 tgcagaaagaaatcacactgccttcaagactcatatattacatcaaccaagactcggaaa 379
Db     17 tggagcgccagagcatcgtgccactgcgcctcatctaccgctcgggcggcgaagacgaaa 76

Qy    380 gcccttatcacgttcttgacacaaaggcaagacaccagcaaaaacataataaggctgtcc 439
Db     77 gtcggcacgacgcgctcgacacgcgggtgcggggcgacctcggtggccggcagttgactc 136

Qy    440 atctggcccaggcaagcttccagattgaagccttcggctccaaattcattcttgacctca 499
Db    137 atgttgaccaagcaagcttccaggttgatgcctttggaacgtcattcattctcgatgtcg 196

Qy    500 tactgaacaatggtttgttgtcttctgattatgtggagattcactacgaaaat---ggga 556
Db    197 tgctaaatcatgatttgctgtcctctgaatacatagagagacacattgaacatggaggca 256

Qy    557 aaccacagtactctaagggtggagagcactgttactaccatggaagcatcagaggcgtca 616
Db    257 agactgtggaagttaaaggaggagagcactgttactaccagggccatatccgaggaaacc 316

Qy    617 aagactccaaggtggctctgtcaacctgcaatggacttcatggcatgtttgaagatgata 676
Db    317 ctgactcatttgttgcattgtcaacatgccacggacttcatgggatgttctatgacggga 376

Qy    677 ccttcgtgtatatgatagagccactagagctggttcatgatgagaaaagcacaggtcgac 736
Db    377 accacacatatctcattgagccagaagaaaatg----acactactcaagaggatttccat 432

Qy    737 cacatataatccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcacta 796
Db    433 tttcattcagtttacaaatccagactgtttgaattttccttggatgatcttccatctgaa 492

Qy    797 tggaaagaggtgaccagtggccctttctctctgaattacagtggttgaaaagaaggaaga 856
Db    493 tttcagcaagtaaacattactccatcaaaatttattttgaagccaagaccaaaaaggagt 552

Qy    857 gagcagtgaatccat-------cacgtggtatatttgaagaaatgaaatatttggaactt 909
Db    553 aaacggcagcttcgtcgatatcctcgtaatgtagaagaagaaaccaaatacattgaactg 612

Qy    910 atgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgcacataccaac 969
Db    613 atgattgtgaatgatcaccttatgtttaaaaaacatcggctttccgttgtacataccaat 672

Qy    970 aactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcagctcaacacc 1029
Db    673 acctatgcgaaatctgtggtgaacatggcagatttaatatataaagaccaacttaagacc 732

Qy   1030 agggttgtcctggtggctgtagagacctggactgagaaggatcagattgacatcaccacc 1089
Db    733 aggatagtattggttgctatggaaacctgggcgactgacaacaagtttgccatatctgaa 792

Qy   1090 aaccctgtgcagatgctccatgagttctcaaaataccg---gcagcgcattaagcagcat 1146
Db    793 aatccattgatcaccctacgtgagtttatgaaatacaggagggattttatcaaagagaaa 852

Qy   1147 gctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagcagtctgagt 1206
Db    853 agtgatgcagttcacctttttttcgggaagtcaatttgagagtagccggagcggggcagct 912

Qy   1207 tactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtatggtcttcca 1266
Db    913 tatattggtgggatttgctcgttgctgaaaggaggaggcgtgaatgaatttgggaaaact 972

Qy   1267 atggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatccaatgggaa 1326
Db    973 gatttaatggctgttacacttgcccagtcattagcccataatattggtattatctcagac 1032

Qy   1327 ccttcta---gcagaaagccaaaatgtgactgcacagaatcctggggtggctgcatcatg 1383
Db   1033 aaaagaaagttagcaagtggtgaatgtaaatgcgaggacacgtggtccgggtgcataatg 1092

Qy   1384 gaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttggagtataga 1443
Db   1093 ggagacactggctattatcttcctaaaaagttcacccagtgtaatattgaagagtatcat 1152

Qy   1444 gacttttacagagaggaggtggagcctgccttttcaacaggccaacaaagctatttgag 1503
Db   1153 gacttcctgaatagtggaggtggtgcctgccttttcaacaaaccttctaagcttcttgat 1212

Qy   1504 cccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgtggttttcat 1563
Db   1213 cctcctgagtgtggcaatggcttcattgaaactggagaggagtgtgattgtggaaccccg 1272

Qy   1564 gtggaatg---------ctatggattatgctgtaagaaatgttccctctccaacggggct 1614
Db   1273 gccgaatgtgtccttgaaggagcagagtgttgtaagaaatgcaccttgactcaagactct 1332

Qy   1615 cactgcagcgacgggccctgctgtaacaatacctcatgtctttttcagccacgagggtat 1674
Db   1333 caatgcagtgacggtctttgctgtaaaaa------gtgcaagtttcagcctatgggcact 1386

Qy   1675 gaatgccgggatgctgtgaacgagtgtgatattactgaatattgtactggagactctggt 1734
Db   1387 gtgtgccgagaagcagtaaatgattgtgatattcgtgaaacgtgctcaggaaattcaagc 1446

Qy   1735 cagtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatcagggccgc 1794
Db   1447 cagtgtgcccctaatattcataaaatggatggatattcatgtgatggtgttcagggaatt 1506
```

```
Qy   1795 tgctacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctggggaacaaag 1854
          ||||     ||    ||||| |||||||| |  || ||  | ||||  |||||   ||||
Db   1507 tgctttggaggaagatgcaaaaccagagatagacaatgcaaatacattttgggggcaaaag 1566

Qy   1855 gctgcagggtctgacaagttctgctatgaaaagctgaatacagaaggcactgagaaggga 1914
          |   |||  || ||||| |  |||||||| || |||||||  ||||  || ||||||||
Db   1567 gtgacagcatcagacaaatattgctatgagaaactgaatattgaaggacggagaagggt 1626

Qy   1915 aactgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtgga 1974
          ||||| ||||| ||   ||||  ||||| ||||||| |||||  | |||| | |||||
Db   1627 aactgtgggaaagacaaagacacatggatacagtgcaacaaacgggatgtgctttgtggt 1686

Qy   1975 ttcttactctgtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatc 2034
          | | |  | |||||||| ||         ||| |  |||| |||  | ||||| |||
     1687 tacctttgtgtaccaatattggcaatatcccaaggcttggagaacacgatggtgaaatc 1746

     2035 attccaacttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagtt 2094
          |   | ||||   |    ||.|||||  |     |  |||||||| ||  |||||
Db   1747 acatctactttagttgtgcagcaaggaagaacattaaactgcagtggtgcgcatgttaag 1806

Qy   2095 ttagatgatgatacggatgtgggctatgtagaagatggaacgccatgtgccccgtctatg 2154
           | || || |||   ||| | |||||||| |||||||| || || ||||| ||    |||
Db   1807 cttgaagaagatgtagatcttggctatgtggaagatgggacaccttgtgctccccaaatg 1866

Qy   2155 atgtgtttagatcggaagtgcctacaaattcaagccctaaatatgagcacctgtccactc 2214
          ||||| ||||| |  | ||| || |   |      |  | ||  | || |  ||      |
Db   1867 atgtgcttagaacacaggtgtcttcctgtggcttctttcaactttagtacttgcttgagc 1926

Qy   2215 gattccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatt 2274
            |    | || |   | || || ||  |||| || || ||||||||    |  ||  |
Db   1927 agtaaagaaggcactatttgctcaggaaatggagtttgcagtaatgagctgaagtgtgtg 1986

Qy   2275 tgtgatttcacctgggcagggacagattg---cagtatccgggatccagttaggaacctt 2331
          ||| |      ||||  |||  | |||||   || |  |    ||    |   |
Db   1987 tgtaacagacactggataggttctgattgcaacacttacttccctcacaatgatgatgca 2046

Qy   2332 caccccccaaggatgaaggacccaagggtcctagtgccaccaatccataataggctcc 2391
           |  |    |    |        ||| |||  |  || |||||| ||||||||||
Db   2047 aagactggtatcactctgtctggcaatggtgttgctggcaccaatacataataggcata 2106

Qy   2392 atcgctggtgccatcctggtagcagctattgtccttgggggcacaggctgggatttaaa 2451
          || |||||  ||||  | ||    ||  |  |  | ||    || |  ||||| | ||||
Db   2107 attgctggcaccatttttagtgctggccctcatattaggaataactgcgtgggggttataaa 2166

Qy   2452 aa 2453
          ||
     2167 aa 2168
```

```
RESULT   6
AAX78437
ID   AAX78437 standard; cDNA to mRNA; 2697 BP.
XX
AC   AAX78437;
XX
DT   26-AUG-1999  (first entry)
XX
DE   Human MDC2-beta cDNA.
XX
KW   Metalloproteinase-like-disintegrin-like cysteine rich protein; human;
KW   MDC2-alpha; MDC2-beta; MDC3; medical; treatment; diagnosis; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             122..2593
FT                   /*tag=  a
FT                   /product= "MDC2-beta"
XX
PN   JP11155574-A.
XX
PD   15-JUN-1999.
XX
PF   01-DEC-1997;   97JP-0330020.
XX
PR   01-DEC-1997;   97JP-0330020.
XX
PA   (EISA ) EISAI CO LTD.
XX
DR   WPI; 1999-398071/34.
DR   P-PSDB; AAY25119.
XX
PT   New protein belonging to MDC gene family  - useful in medical
PT   treatment and diagnosis
XX
PS   Claim 9; Page 10-13; 17pp; Japanese.
XX
CC   This invention describes novel human MDC2-alpha, MDC2-beta and MDC3
CC   proteins and their encoding nucleic acids. The products of the invention
CC   are useful in medical treatment and diagnosis.
XX
SQ   Sequence 2697 BP; 795 A; 529 C; 699 G; 674 T; 0 other;
```

```
  Query Match             23.5%;  Score 586.8;  DB 20;  Length 2697;
  Best Local Similarity   57.2%;  Pred. No. 5.6e-150;
  Matches 1237;  Conservative    0;  Mismatches  887;  Indels   38;  Gaps    8;

Qy    320 tgcagaaagaaatcacactgccttcaagactcatatattacatcaaccaagactcggaaa 379
          || ||    | | ||   ||||     | ||||| ||   |     |   ||   ||||
Db    255 tggagcgccagagcatcgtgccactgcgcctcatctaccgctcgggcggcgaagacgaaa 314

Qy    380 gcccttatcacgttcttgacacaaaggcaagacaccagcaaaaacataataaggctgtcc 439
          | |   |  |||  || |||||   ||   |   | | |           ||      |
Db    315 gtcggcacgacgcgctcgacacgcgggtgcggggcgacctcggtggcccgcagttgactc 374

Qy    440 atctggcccaggcaagcttccagattgaagccttcggctccaaattcattcttgacctca 499
          || | | ||| ||||||||||||| |||| ||||| ||   |   |||||||| ||  ||
Db    375 atgttgaccaagcaagcttccaggttgatgccctttggaacgtcattcattctcgatgtcg 434

Qy    500 tactgaacaatggtttgttgtcttctgattatgtggagattcactacgaaaat---ggga 556
          | || ||  ||| |||| |||| ||||| ||  | ||||  |||   ||| ||   || |
Db    435 tgctaaatcatgatttgctgtcctctgaatacatagagagacacattgaacatggaggca 494

Qy    557 aaccacagtactctaagggtggagagcactgttactaccatggaagcatcagaggcgtca 616
          |  |   | |   ||| || ||||||||||||||||||||| ||    ||| ||||   |
Db    495 agactgtggaagttaaaggaggagagcactgttactaccagggccatatccgaggaaacc 554

Qy    617 aagactccaaggtggctctgtcaacctgcaatggacttcatggcatgtttgaagatgata 676
            |||||    || ||  |||||||| ||| | ||||||||||| |||||  | || |  |
Db    555 ctgactcatttgttgcattgtcaacatgccacggacttcatgggatgttctatgacggga 614

Qy    677 ccttcgtgtatatgatagagccactagagctggttcatgatgagaaaagcacaggtcgac 736
            |  |   ||| | || ||||||  |||    |      |  |   |||  |  || |
Db    615 accacacatatctcattgagccagaagaaaatg----acactactcaagaggatttccat 670

Qy    737 cacatataatccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcacta 796
              | |      ||| |  | | |      ||      |        ||   |
Db    671 tttcattcagtttacaaatccagactgtttgaattttccttggatgatcttccatctgaa 730

Qy    797 tggaaagaggtgaccagtggccctttctctctgaattacagtggttgaaaagaaggaaga 856
          |   |  | || | || |   || |      | | ||  |      ||  | || | ||
Db    731 tttcagcaagtaaacattactccatcaaaatttattttgaagccaagaccaaaaaggagt 790

Qy    857 gagcagtgaatccat-------cacgtggtatatttgaagaaatgaaatatttggaactt 909
            | | |    | | |       | |||  | ||   |||||||  |||||  | |||||
Db    791 aaacggcagcttcgtcgatatcctcgtaatgtagaagaagaaaccaaatacattgaactg 850

Qy    910 atgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgcacataccaac 969
          ||||||||| |||||||||   | || ||| || |||||   |||   || |||||||||
Db    851 atgattgtgaatgatcaccttatgtttaaaaaacatcggctttccgttgtacataccaat 910

Qy    970 aactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcagctcaacacc 1029
          | || ||| || || |||||  ||| | |  ||||  || || || || || || ||| |||
Db    911 acctatgcgaaatctgtggtgaacatggcagatttaatatataaagaccaacttaagacc 970
```

```
Qy   1030 agggttgtcctggtggctgtagagacctggactgagaaggatcagattgacatcaccacc 1089
          ||| | ||  |||| ||| | || |||||| |      |  |  || ||| |||  |
Db    971 aggatagtattggttgctatggaaacctgggcgactgacaacaagtttgccatatctgaa 1030

Qy   1090 aaccctgtgcagatgctccatgagttctcaaaataccg---gcagcgcattaagcagcat 1146
          || ||  ||   |  || | ||||||    |||||| |   | |    || ||  || |
Db   1031 aatccattgatcaccctacgtgagtttatgaaatacaggagggattttatcaaagagaaa 1090

Qy   1147 gctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagcagtctgagt 1206
           ||||||  || |||||  | ||| |      |||| |   ||   | ||| |      |
Db   1091 agtgatgcagttcacctttttcgggaagtcaatttgagagtagccggagcggggcagct 1150

Qy   1207 tactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtatggtcttcca 1266
          ||  |||| ||  | || ||       | |||||  || |||||||| | |||     |
Db   1151 tatattggtgggatttgctcgttgctgaaaggaggaggcgtgaatgaatttgggaaaact 1210

Qy   1267 atggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatccaatgggaa 1326
               | ||||      | |  | |||    | || || ||  |||| ||    |  ||
Db   1211 gatttaatggctgttacacttgcccagtcattagcccataatattggtattatctcagac 1270

Qy   1327 ccttcta---gcagaaagccaaaatgtgactgcacagaatcctggggtggctgcatcatg 1383
                |       ||      ||||| | |||   ||  | |||   || ||||| |||
Db   1271 aaaagaaagttagcaagtggtgaatgtaaatgcgaggacacgtggtccgggtgcataatg 1330

Qy   1384 gaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttggagtataga 1443
          |  || || ||    |  | | ||  ||| ||  |  |||| |  |||   ||||||
Db   1331 ggagacactggctattatcttcctaaaaagttcacccagtgtaatattgaagagtatcat 1390

Qy   1444 gacttttacagagaggaggtggagcctgccttttcaacaggccaacaaagctatttgag 1503
          |||||  |  | || |||||||| ||||||||||||||||  ||  | |||||  ||||
Db   1391 gacttcctgaatagtggaggtggtgcctgccttttcaacaaaccttctaagcttcttgat 1450

Qy   1504 cccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgtggttttcat 1563
          ||  | || ||||| ||||| | | | ||| |||| ||||||||||||||||    |
Db   1451 cctcctgagtgtggcaatggcttcattgaaactggagaggagtgtgattgtggaaccccg 1510

Qy   1564 gtggaatg---------ctatggattatgctgtaagaaatgttccctctccaacggggct 1614
          |  |||||             | |   || ||||||||||  || |  |  | |   ||
Db   1511 gccgaatgtgtccttgaaggagcagagtgttgtaagaaatgcaccttgactcaagactct 1570

Qy   1615 cactgcagcgacgggccctgctgtaacaatacctcatgtctttttcagccacgagggtat 1674
          || ||||| ||||| |  |||||||| ||       ||    ||||||||    ||   |
Db   1571 caatgcagtgacggtctttgctgtaaaaa------gtgcaagtttcagcctatgggcact 1624

Qy   1675 gaatgccgggatgctgtgaacgagtgtgatattactgaatattgtactggagactctggt 1734
          |  ||||| || || || || || |||||||||  |||| ||  | ||| | ||  |
Db   1625 gtgtgccgagaagcagtaaatgattgtgatattcgtgaaacgtgctcaggaaattcaagc 1684

Qy   1735 cagtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatcagggccgc 1794
          |||||  | || ||| |||||||    || |||||| |||  ||     ||||||
Db   1685 cagtgtgcccctaatattcataaaatggatggatattcatgtgatggtgttcagggaatt 1744

Qy   1795 tgctacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctggggaacaaag 1854
          ||||     ||    ||||| |||||||| |  || ||  | ||||| |||||   ||||
Db   1745 tgctttggaggaagatgcaaaaccagagatagacaatgcaaatacatttgggggcaaaag 1804

Qy   1855 gctgcagggtctgacaagttctgctatgaaaagctgaatacagaaggcactgagaaggga 1914
          |   |||  || ||||| |  |||||||| || |||||||  |||||  || |||||||
Db   1805 gtgacagcatcagacaaatattgctatgagaaactgaatattgaagggacggagaagggt 1864

Qy   1915 aactgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtgga 1974
          ||||| ||||| ||   ||||   ||||| |||||| |||||  ||||||  | |||||
Db   1865 aactgtgggaaagacaaagacacatggatacagtgcaacaaacgggatgtgctttgtggt 1924

Qy   1975 ttcttactctgtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatc 2034
          | | |  | |||||||||  ||         ||| |  |||| ||||  | |||||  |||
Db   1925 taccttttgtgtaccaatattggcaatatcccaaggcttggagaactcgatggtgaaatc 1984

Qy   2035 attccaacttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagtt 2094
          |   | ||||   |    || |||||  |     |  ||||||||||||  |||||
Db   1985 acatctactttagttgtgcagcaaggaagaacattaaactgcagtggtgggcatgttaag 2044

Qy   2095 ttagatgatgatacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatg 2154
```

```
               | || || |||   ||| | |||||||| |||||||| || || ||||| ||    |||
Db   2045 cttgaagaagatgtagatcttggctatgtggaagatgggacaccttgtggtccccaaatg 2104

Qy   2155 atgtgtttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactc 2214
          ||||| ||||| |  | ||| || |   |      |  | ||  | || |  ||      |
Db   2105 atgtgcttagaacacaggtgtcttcctgtggcttctttcaactttagtacttgcttgagc 2164

Qy   2215 gattccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatt 2274
            |    | || |  | || || ||  |||| || || ||||||||    |  ||  |
Db   2165 agtaaagaaggcactatttgctcaggaaatggagtttgcagtaatgagctgaagtgtgtg 2224

Qy   2275 tgtgatttcacctgggcagggacagattg---cagtatccgggatccagttaggaacctt 2331
          ||| |      ||||  |||  | |||||   || |  |     ||    |   |
Db   2225 tgtaacagacactggataggttctgattgcaacacttacttccctcacaatgatgatgca 2284

Qy   2332 caccccccaaggatgaaggacccaagggtcctagtgccaccaatctcataataggctcc 2391
           |  |    |    |         ||| |||  |  || |||||||  ||||||||||
Db   2285 aagactggtatcactctgtctggcaatggtgttgctggcaccaatatcataataggcata 2344

Qy   2392 atcgctggtgccatcctggtagcagctattgtccttgggggcacaggctggggatttaaa 2451
          || |||||  ||||  | ||    ||  |  |  | ||    || |  ||||| | ||||
Db   2345 attgctggcaccattttagtgctggccctcatattaggaataactgcgtggggttataaa 2404

Qy   2452 aa 2453
          ||
Db   2405 aa 2406
```

RESULT 7
AAX78436
ID AAX78436 standard; cDNA to mRNA; 2805 BP.
XX
AC AAX78436;
XX
DT 26-AUG-1999 (first entry)
XX
DE Human MDC2-alpha cDNA.
XX
KW Metalloproteinase-like-disintegrin-like cysteine rich protein; human;
KW MDC2-alpha; MDC2-beta; MDC3; medical; treatment; diagnosis; ds.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	CDS	122..2701
FT		/*tag= a
FT		/product= "MDC-2 alpha"

XX
PN JP11155574-A.
XX
PD 15-JUN-1999.
XX
PF 01-DEC-1997; 97JP-0330020.
XX
PR 01-DEC-1997; 97JP-0330020.
XX
PA (EISA) EISAI CO LTD.
XX
DR WPI; 1999-398071/34.
DR P-PSDB; AAY25118.
XX
PT New protein belonging to MDC gene family - useful in medical
PT treatment and diagnosis
XX
PS Claim 1; Page 6-10; 17pp; Japanese.
XX
CC This invention describes novel human MDC2-alpha, MDC2-beta and MDC3
CC proteins and their encoding nucleic acids. The products of the invention
CC are useful in medical treatment and diagnosis.
XX
SQ Sequence 2805 BP; 829 A; 553 C; 723 G; 700 T; 0 other;

```
  Query Match             23.5%;  Score 586.8;  DB 20;  Length 2805;
  Best Local Similarity   57.2%;  Pred. No. 5.7e-150;
  Matches 1237;  Conservative    0;  Mismatches  887;  Indels   38;  Gaps    8;

Qy     320 tgcagaaagaaatcacactgccttcaagactcatatattacatcaaccaagactcggaaa 379
           || ||    | | ||   ||||     | ||||| ||   |      |   ||    ||||
Db     255 tggagcgccagagcatcgtgccactgcgcctcatctaccgctcgggcggcgaagacgaaa 314

Qy     380 gcccttatcacgttcttgacacaaaggcaagacaccagcaaaaacataataaggctgtcc 439
           | |    |  |||  || |||||   ||   |   | | |              ||      |
Db     315 gtcggcacgacgcgctcgacacgcgggtgcggggcgacctcggtggcccgcagttgactc 374

Qy     440 atctggcccaggcaagcttccagattgaagccttcggctccaaattcattcttgacctca 499
           || | | ||| |||||||||||| |||| ||||| ||   |   |||||||| ||  ||
       375 atgttgaccaagcaagcttccaggttgatgccttggaacgtcattcattctcgatgtcg 434

       500 tactgaacaatggtttgttgtcttctgattatgtggagattcactacgaaaat---ggga 556
           | || ||  ||| |||| |||| ||||| ||  | ||||  |||   ||| ||   || |
Db     435 tgctaaatcatgatttgctgtcctctgaatacatagagagacacattgaacatggaggca 494

Qy     557 aaccacagtactctaagggtggagagcactgttactaccatggaagcatcagaggcgtca 616
           |  |   | |   ||| || ||||||||||||||||||||| ||    ||| ||||   |
Db     495 agactgtggaagttaaaggaggagagcactgttactaccagggccatatccgaggaaacc 554

Qy     617 aagactccaaggtggctctgtcaacctgcaatggacttcatggcatgtttgaagatgata 676
             |||||    || ||  |||||| ||| | ||||||||||| |||||  | || |  |
Db     555 ctgactcatttgttgcattgtcaacatgccacggacttcatgggatgttctatgacggga 614

Qy     677 ccttcgtgtatatgatagagccactagagctggttcatgatgagaaaagcacaggtcgac 736
           |  |   ||| | || ||||||  |||    |      |  |   |||   |  || |
Db     615 accacacatatctcattgagccagaagaaaatg----acactactcaagaggatttccat 670

Qy     737 cacatataatccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcacta 796
           |  |       ||| |  | | |       ||       |           ||   |
Db     671 tttcattcagtttacaaatccagactgtttgaattttccttggatgatcttccatctgaa 730

Qy     797 tggaaagaggtgaccagtggccctttctctctgaattacagtggttgaaaagaaggaaga 856
           |   |  | || | || |   || |      | | ||  |      ||  | || | ||
Db     731 tttcagcaagtaaacattactccatcaaaatttattttgaagccaagaccaaaaaggagt 790

Qy     857 gagcagtgaatccat-------cacgtggtatatttgaagaaatgaaatatttggaacttt 909
           | | |    | | |       | |||  | ||   |||||||  |||||  | |||||
Db     791 aaacggcagcttcgtcgatatcctcgtaatgtagaagaagaaaccaaatacattgaactg 850

Qy     910 atgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgcacataccaac 969
           |||||||| |||||||||   | || ||| || |||||   |||   || |||||||||
Db     851 atgattgtgaatgatcaccttatgtttaaaaaacatcggctttccgttgtacataccaat 910

       970 aactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcagctcaacacc 1029
           | || ||| || || ||||| ||| | |  |||| || || || || || || || |||
Db     911 acctatgcgaaatctgtggtgaacatggcagatttaatatataaagaccaacttaagacc 970

Qy    1030 agggttgtcctggtggctgtagagacctggactgagaaggatcagattgacatcaccacc 1089
           ||| | ||  |||| ||| | || ||||||  |     |  | || ||| |||  |
Db     971 aggatagtattggttgctatggaaacctgggcgactgacaacaagttgccatatctgaa 1030

Qy    1090 aaccctgtgcagatgctccatgagtttctcaaaataccg---gcagccgcattaagcagcat 1146
           || ||  ||    |  || | ||||||    ||||||  |   | |    || ||  || |
Db    1031 aatccattgatcaccctacgtgagtttatgaaatacaggagggatttatcaaagagaaa 1090

Qy    1147 gctgatgctgtgcacctcatctcgcgggtgacatttcactataagaaagcagtctgagt 1206
             |||||| || |||||  | ||| |      |||| |   ||   | ||| |      |
Db    1091 agtgatgcagttcacctttttcgggaagtcaatttgagagtagccggagcggggcagct 1150

Qy    1207 tactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtatggtcttcca 1266
           ||  |||| ||  | || ||       | |||||  || |||||||  | |||     |
Db    1151 tatattggtgggatttgctcgttgctgaaaggaggaggcgtgaatgaatttgggaaaact 1210

Qy    1267 atggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatccaatgggaa 1326
             | ||||      | |  | |||    | || || ||  |||  ||     |  ||
Db    1211 gatttaatggctgttacacttgcccagtcattagcccataatattggtattatctcagac 1270

Qy    1327 ccttcta---gcagaaagccaaaatgtgactgcacagaatcctgggggtggctgcatcatg 1383

                |      ||      ||||| | |||   ||  | |||   || ||||| |||
Db    1271 aaaagaaagttagcaagtggtgaatgtaaatgcgaggacacgtggtccgggtgcataatg 1330

Qy    1384 gaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttggagtataga 1443
           |  || || ||     |  | | ||  ||| ||  |  |||| |  |||   ||||||
Db    1331 ggagacactggctattatcttcctaaaaagttcacccagtgtaatattgaagagtatcat 1390

Qy    1444 gacttttacagagaggaggtggagcctgccttttcaacaggccaacaaagctatttgag 1503
           |||||  |  | || |||||||| ||||||||||||||||  ||  | |||||  ||||
Db    1391 gacttcctgaatagtggaggtggtgcctgccttttcaacaaaccttctaagcttcttgat 1450

Qy    1504 cccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgtggttttcat 1563
           ||  | || ||||| ||||| | | | ||| |||| ||||||||||||||||||    |
Db    1451 cctcctgagtgtggcaatggcttcattgaaactggagaggagtgtgattgtggaaccccg 1510

Qy    1564 gtggaatg---------ctatggattatgctgtaagaaatgttccctctccaacggggct 1614
           |  |||||              | |   || ||||||||||  || |  |  | |   ||
Db    1511 gccgaatgtgtccttgaaggagcagagtgttgtaagaaatgcaccttgactcaagactct 1570

Qy    1615 cactgcagcgacgggccctgctgtaacaatacctcatgtctttttcagccacgagggtat 1674
           || ||||| ||||| |  |||||||| ||       ||    ||||||||   ||  |
Db    1571 caatgcagtgacggtctttgctgtaaaaa------gtgcaagtttcagcctatgggcact 1624

Qy    1675 gaatgccgggatgctgtgaacgagtgtgatattactgaatattgtactggagactctggt 1734
           |  ||||| || || || || || |||||||||  ||||   ||  | ||| | ||  |
Db    1625 gtgtgccgagaagcagtaaatgattgtgatattcgtgaaacgtgctcaggaaattcaagc 1684

Qy    1735 cagtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatcagggccgc 1794
           |||||  | || ||| |||||||    || |||||| |||  ||     ||||||
Db    1685 cagtgtgcccctaatattcataaaatggatggatattcatgtgatggtgttcagggaatt 1744

Qy    1795 tgctacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctggggaacaaag 1854
           ||||     ||    ||||| |||||||| |  || ||  | ||||| |||||   ||||
Db    1745 tgctttggaggaagatgcaaaaccagagatagacaatgcaaatacatttgggggcaaaag 1804

Qy    1855 gctgcagggtctgacaagttctgctatgaaaagctgaatacagaaggcactgagaaggga 1914
           |   |||  || |||||  |  |||||||| || |||||||  ||||| || ||||||||
Db    1805 gtgacagcatcagacaaatattgctatgagaaactgaatattgaagggacggagaagggt 1864

Qy    1915 aactgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtgga 1974
           ||||| ||||| ||   ||||  ||||| |||||||  |||||  ||||||  | |||||
Db    1865 aactgtgggaaagacaaagacacatggatacagtgcaacaaacgggatgtgctttgtggt 1924

Qy    1975 ttcttactctgtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatc 2034
           | | |  | |||||||||  ||         ||| |  ||||  ||||  | ||||| |||
Db    1925 taccttttgtgtaccaatattggcaatatcccaaggcttggagaactcgatggtgaaatc 1984

Qy    2035 attccaacttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagtt 2094
           |   | ||||   |    || |||||  |     | ||||||||||||  |||||
Db    1985 acatctacttttagttgtgcagcaaggaagaacattaaactgcagtggtgggcatgttaag 2044

Qy    2095 ttagatgatgatacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatg 2154
           | || || |||   ||| | |||||||| |||||||| || || |||||  ||    |||
Db    2045 cttgaagaagatgtagatcttggctatgtggaagatgggacaccttgtggtccccaaatg 2104

Qy    2155 atgtgtttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactc 2214
           ||||| ||||| |  | ||| || |   |     |  | ||  | || |  ||      |
Db    2105 atgtgcttagaacacaggtgtcttcctgtggcttctttcaactttagtacttgcttgagc 2164

Qy    2215 gattccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatt 2274
             |    | || |   | || || ||  |||| || || ||||||||    |  ||  |
Db    2165 agtaaagaaggcactatttgctcaggaaatggagtttgcagtaatgagctgaagtgtgtg 2224

Qy    2275 tgtgatttcacctgggcagggacagattg---cagtatccgggatccagttaggaaccttt 2331
           ||| |      |||| |||  | |||||   || |  |     ||   |    |
Db    2225 tgtaacagacactggataggttctgattgcaacacttacttccctcacaatgatgatgca 2284

Qy    2332 caccccccaaggatgaaggacccaagggtcctagtgccaccaatctcataataggctcc 2391
           |  |    |   |        ||| |||  |  || |||||||  ||||||||||
Db    2285 aagactggtatcactctgtctggcaatggtgttgctggcaccaatatcataataggcata 2344

Qy    2392 atcgctggtgccatcctggtagcagctattgtccttgggggcacaggctggggatttaaa 2451
           || |||||  ||||  | ||    ||  |  |  | ||    || |  |||||  | ||||
```

```
Db    2345 attgctggcaccattttagtgctggccctcatattaggaataactgcgtggggttataaa 2404

Qy    2452 aa 2453
           ||
Db    2405 aa 2406


RESULT   8
AAQ76122
ID   AAQ76122 standard; cDNA to mRNA; 3183 BP.
XX
AC   AAQ76122;
XX
DT   01-AUG-1995  (first entry)
XX
DE   Human fetal brain MDC cDNA.
XX
KW   MDC protein; breast cancer; mamma carcinoma; ovary cancer;
KW   chromosome-17; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..2310
FT                   /*tag= a
FT   3'UTR           2311..3183
FT                   /*tag= b
XX
PN   EP633268-A.
XX
PD   11-JAN-1995.
XX
PF   13-MAY-1994;   94EP-0107487.
XX
PR   14-MAY-1993;   93JP-0136602.
PR   22-SEP-1993;   93JP-0257455.
PR   23-FEB-1994;   94JP-0049904.
PR   12-APR-1994;   94JP-0073328.
XX
PA   (CANC-) CANCER INST.
PA   (EISA ) EISAI CO LTD.
XX
PI   Emi M,  Nakamura Y;
XX
DR   WPI; 1995-038478/06.
DR   P-PSDB; AAR75352.
XX
PT   Novel MDC protein and DNA encoding it - used to develop prods.
PT   for the study, diagnosis and therapy of cancers, partic. breast
PT   and ovarian cancer
XX
PS   Disclosure; Page 94-99; 123pp; English.
XX
CC   A detailed map of human chromosome-17 was constructed to analyze the
CC   chromosome in breast and ovarian cancer tissues.  2 Novel cDNAs
CC   (given in AAQ76120-21) were obtained from fetal cerebellum that
CC   encoded novel MDC proteins (AAR67758-59, respectively).  5'-RACE and
CC   RT-PCR revealed the sequence given in AAQ76122, encoding the MDC given
CC   in AAR75352.  Sequences common for these clones are given in AAQ76119
CC   and AAR66757, respectively.  A genomic DNA sequence (AAQ76124) for MDC
CC   was obtained from cosmid clone cCI17-904.
XX
SQ   Sequence 3183 BP; 645 A; 983 C; 978 G; 577 T; 0 other;


  Query Match             23.0%;  Score 575;  DB 16;  Length 3183;
  Best Local Similarity   57.5%;  Pred. No. 1e-146;
  Matches 1226;  Conservative    0;  Mismatches  860;  Indels   45;  Gaps    9;

Qy     376 gaaagcccttatcacgttcttgacacaaaggcaagacaccagcaaaaacataataaggct 435
           ||   ||   | ||    || |||||||  ||   | ||  ||| |  |        | ||
Db     181 gaggtccgaaagcagcagctggacacaagggtccgccaggagccaccaggggcccgcct 240

Qy     436 gtccatctggcccaggcaagcttccagattgaagccttcggctccaaattcattcttgac 495
           ||||||||||||||||  || |||   ||   |||||||  ||| || ||||  || |||
Db     241 gtccatctggcccaggtgagtttcgtcatcccagccttcaactcaaacttcaccctggac 300

Qy     496 ctcatactgaacaatggtttgttgtcttctgattatgtggagattcactac------gaa 549
           ||    |||||| |      |  | || ||  | || ||||||   |||| |      ||
Db     301 ctggagctgaaccaccacctcctctcctcgcaatacgtggagcgccacttcagccgggag 360

Qy     550 aatgggaaaccacagtactctaagggtggagagcactgttactaccatggaagcatcaga 609
               |  |  ||   | |   || |||||| ||||| |||||||| || |   || |
Db     361 gggacaacccagcacagcaccggggctggagaccactgctactaccaggggaagctccgg 420

Qy     610 ggcgtcaaagactccaaggtggctctgtcaacctgcaatggacttcatggcatgtttgaa 669
           ||   |    |||||   |  || || || |||||| | || || |||||  | ||
Db     421 gggaacccgcactccttcgccgccctctccacctgccaggggctgcatggggtcttctct 480

Qy     670 gatgataccttcgtgtatatgatagagccactagagctggttcatgatgagaaaagcac- 728
           ||||  | |||    || ||  | ||||| | |||| ||| |     |  |  |  | |
Db     481 gatgggaacttgacttacatcgtggagccccaagaggtggctggaccttggggagcccct 540

Qy     729 -----aggtcgaccacatataatccagaaaaccttggcaggacagtattctaagcaaatg 783
                |   | || ||  | ||  |    |||          |   || |    |
Db     541 cagggacccctcccacctcatttaccggaccccctctcctcccagatcccctcggatgc 600

Qy     784 aagaatctcactatggaaagaggtgaccagtggccctttctctctgaattacagtggttg 843
           | | | |               | ||  |     |  |      ||  |   | |  | |
Db     601 agggaaccaggctgcctgtttgctgtgcctgcccagtcggctcctccaaaccggccgagg 660

Qy     844 aaaagaaggaagagagcagtgaatccatcacgtggtatatttga---agaaatgaaatat 900
              |||||||| ||    ||    |     |    || | |  |    ||||  || |||
Db     661 ctgagaaggaaaaggcaggtccgccggggccaccctacagtgcacagtgaaaccaagtat 720

Qy     901 ttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgca 960
           ||||  || ||  |  | || || ||| |   ||   || ||   ||    ||    |
Db     721 gtggagctaattgtgatcaacgaccaccagctgttcgagcagatgcgacagtcggtggtc 780

Qy     961 cataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcag 1020
           |  |||| |||||||||| |||||||||||| ||||| |  |||   || ||||||||||||
Db     781 ctcaccagcaactttgccaagtccgtggtgaacctggccgatgtgatatacaaggagcag 840

Qy     1021 ctcaacaccagggttgtcctggtggctgtagagacctggactgagaaggatcagattgac 1080
            |||||||  |  | ||||||||  ||  | ||  || ||| | ||   |||  ||||  |
Db     841 ctcaacactcgcatcgtcctggttgccatggaaacatgggcagatggggacaagatccag 900

Qy     1081 atcaccaccaaccctgtgcagatgctccatgagttctcaaaatac---cggcagcgcatt 1137
            |        |||   || |||  ||      | ||      |||   ||| || |  |
Db     901 gtgcaggatgacctcctggagaccctggcccggctcatggtctaccgacgggagggtctg 960

Qy     1138 aagcagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagc 1197
              |||    | ||||   |||||| ||||| |   ||| || ||   | ||| |||
Db     961 cctgagcccagtaatgccacccacctcttctcgggcaggaccttccagagcacgagcagc 1020

Qy     1198 agtctgagttactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtat 1257
            |       ||| | || ||  | || || |   |    || |  |||||||| |||||
Db     1021 ggggcagcctacgtgggggggcatatgctccctgtcccatggcgggggtgtgaacgagtac 1080

Qy     1258 ggtcttccaatggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatc 1317
            ||         |||  ||||        |  | ||||  ||||  || ||||| || ||
Db     1081 ggcaacatgggggcgatggccgtgacccttgcccagacgctgggacagaacctgggcatg 1140

Qy     1318 caatgggaaccttctagcagaaagccaa---aatgtgactgcacagaatcctggggtggc 1374
              ||| |        | ||    | ||  | ||  |  ||||    ||||   |||
Db     1141 atgtggaacaaacaccggagctcggcaggggactgcaagtgtccagacatctggctgggc 1200

Qy     1375 tgcatcatggaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttg 1434
            ||||||||||||||| || ||| | | ||   | || || || ||    ||||||||
Db     1201 tgcatcatggaggacactgggttctacctgccccgcaagttctctcgctgcagcatcgac 1260

Qy     1435 gagtatagagacttttttacagagaggaggtggagcctgccttttcaacaggccaacaaag 1494
            ||||| |   | ||| | |||   || |||||   |||||| ||||||| |||    |||
Db     1261 gagtacaaccagtttctgcaggagggtggtggcagctgcctcttcaacaagcccctcaag 1320

Qy     1495 ctatttgagcccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgt 1554
```

```
Db   1321 ctcctggaccccccagagtgcgggaacggcttcgtggaggcaggggcggagtgcgactgc 1380

Qy   1555 ggttttcatgtggaatgc---------tatggattatgctgtaagaaatgttccctctcc 1605

Db   1381 ggctcggtgcaggagtgcagccgcgcaggtggcaactgctgcaagaaatgcaccctgact 1440

Qy   1606 aacggggctcactgcagcgacgggccctgctgtaacaatacctcatgtctttttcagcca 1665

Db   1441 cacgacgccatgtgcagcgacgggctctgctg------tcgccgctgcaagtacgaacca 1494

Qy   1666 cgagggtatgaatgccgggatgctgtgaacgagtgtgatattactgaatattgtactgga 1725

Db   1495 cggggtgtgtcctgccgagaggccgtgaacgagtgcgacatcgcggagacctgcaccggg 1554

Qy   1726 gactctggtcagtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaat 1785

Db   1555 gactctagccagtgcccgcctaacctgcacaagctggacggttactactgtgaccatgag 1614

Qy   1786 cagggccgctgctacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctgg 1845

Db   1615 cagggccgctgctacggaggtcgctgcaaaacccgggaccggcagtgccaggttctttgg 1674

Qy   1846 ggaacaaaggctgcagggtctgacaagttctgctatgaaaagctgaatacagaaggcact 1905

Db   1675 ggccatgcggctg------ctgatcgcttctgctacgagaagctgaatgtggaggggacg 1728

Qy   1906 gagaagggaaactgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtg 1965

Db   1729 gagcgtgggagctgtgggcgcaagggatccggctgggtccagtgcagtaagcaggacgtg 1788

Qy   1966 ttctgtggattcttactctgtaccaatcttactcgagctccacgtattggtcaacttcag 2025

Db   1789 ctgtgtggcttcctcctctgtgtcaacatctctggagctcctcggctagggacctggtg 1848

Qy   2026 ggtgagatcattccaacttccttctaccatcaaggccgggtgattgactgcagtggtgcc 2085

Db   1849 ggagacatcagtagtgtcaccttctaccaccagggcaaggagctggactgcaggggaggc 1908

Qy   2086 catgtagttttagatgatgatacggatgtgggctatgtagaagatggaacgccatgtggc 2145

Db   1909 cacgtgcagctggcggacggctctgacctgagctatgtggaggatggcacagcctgcggg 1968

Qy   2146 ccgtctatgatgtgtttagatcggaagtgcctacaaattcaagccctaaatatgagcagc 2205

     1969 cctaacatgttgtgcctggaccatcgctgcctgccagcttctgccttcaacttcagcacc 2028

     2206 tgtccactcgattccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagcc 2265

Db   2029 tgccccggcagtggggagcgccggatttgctcccaccacggggtctgcagcaatgaaggg 2088

Qy   2266 acctgcatttgtgatttcacctgggcagggacagattgcagtatccgggatccagttagg 2325

Db   2089 aagtgcatctgtcagccagactggacaggcaaagactgcagtatccaaacccctgccc 2148

Qy   2326 aaccttcaccccccaaggat---gaaggacccaagggtcctagtgccaccaatctcata 2382

Db   2149 acgtccccacccacgggggagacggagagatataaaggtcccagcggcaccaacatcatc 2208

Qy   2383 ataggctccatcgctggtgccatcctggtagcagctattgtccttgggggcacaggctgg 2442

Db   2209 attggctccatcgctggggctgtcctggttgcagccatcgtcctgggcggcacgggctgg 2268

Qy   2443 ggatttaaaaatgtcaagaagagaaggttcg 2473

Db   2269 ggatttaaaaacattcgccgaggaaggtccg 2299


RESULT   9
AAQ76121
ID   AAQ76121 standard; cDNA to mRNA; 2913 BP.
XX
AC   AAQ76121;
XX
DT   01-AUG-1995  (first entry)
XX
DE   Human fetal cerebellum MDC cDNA.
XX
KW   MDC protein; breast cancer; mamma carcinoma; ovary cancer;
KW   chromosome-17; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..27
FT                   /*tag=  a
FT   CDS             28..2340
FT                   /*tag=  b
FT   3'UTR           2038..2913
FT                   /*tag=  c
XX
PN   EP633268-A.
XX
PD   11-JAN-1995.
XX
PF   13-MAY-1994;   94EP-0107487.
XX
PR   14-MAY-1993;   93JP-0136602.
PR   22-SEP-1993;   93JP-0257455.
PR   23-FEB-1994;   94JP-0049904.
PR   12-APR-1994;   94JP-0073328.
XX
PA   (CANC-) CANCER INST.
PA   (EISA ) EISAI CO LTD.
XX
PI   Emi M,  Nakamura Y;
XX
DR   WPI; 1995-038478/06.
DR   P-PSDB; AAR67759.
XX
PT   Novel MDC protein and DNA encoding it - used to develop prods.
PT   for the study, diagnosis and therapy of cancers, partic. breast
PT   and ovarian cancer
XX
PS   Disclosure; Page 90-94; 123pp; English.
XX
CC   A detailed map of human chromosome-17 was constructed to analyze the
CC   chromosome in breast and ovarian cancer tissues.  2 Novel cDNAs
CC   (given in AAQ76120-21) were obtained from fetal cerebellum that
CC   encoded novel MDC proteins (AAR67758-59, respectively).  5'-RACE and
CC   RT-PCR revealed the sequence given in AAQ76122, encoding the MDC given
CC   in AAR75352.  Sequences common for these clones are given in AAQ76119
CC   and AAR66757, respectively.  A genomic DNA sequence (AAQ76124) for MDC
CC   was obtained from cosmid clone cCI17-904.
XX
SQ   Sequence 2913 BP; 604 A; 888 C; 885 G; 536 T; 0 other;


  Query Match             21.2%;  Score 530;  DB 16;  Length 2913;
  Best Local Similarity   57.8%;  Pred. No. 1.9e-134;
  Matches 1114;  Conservative    0;  Mismatches  775;  Indels   39;  Gaps    8;

Qy    573 gggtggagagcactgttactaccatggaagcatcagaggcgtcaaagactccaaggtggc 632

Db    114 ggctggagaccactgctactaccaggggaagctccgggggaacccgcactccttcgccgc 173

Qy    633 tctgtcaacctgcaatggacttcatggcatgtttgaagatgataccttcgtgtatatgat 692

Db    174 cctctccacctgccaggggctgcatggggtcttctctgatgggaacttgacttacatcgt 233

Qy    693 agagccactagagctggttcatgatgagaaaagcac------aggtcgaccacatataat 746

Db    234 ggagccccaagaggtggctggaccttggggagcccctcagggacccctttccccacctcat 293

Qy    747 ccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcactatggaaagagg 806
```

```
Db     294 ttaccggacccctctcctcccagatcccctcggatgcagggaaccaggctgcctgtttgc 353

Qy     807 tgaccagtggccctttctctctgaattacagtggttgaaaagaaggaagagagcagtgaa 866
           ||  |     |  |       ||  |   | |  |  |   |||||||| ||    ||
Db     354 tgtgcctgcccagtcggctcctccaaaccggccgaggctgagaaggaaaaggcaggtccg 413

Qy     867 tccatcacgtggtatatttga---agaaatgaaatatttggaacttatgattgttaatga 923
           |     |    || | |  |   .||||  || ||| |||| || ||  |  | || ||
Db     414 ccggggccaccctacagtgcacagtgaaaccaagtatgtggagctaattgtgatcaacga 473

Qy     924 tcacaaaacgtataagaagcatcgctcttctcatgcacataccaacaactttgcaaagtc 983
           ||| |   ||   || ||   ||    ||    |  |  |||| |||||||| |||||
Db     474 ccaccagctgttcgagcagatgcgacagtcggtggtcctcaccagcaactttgccaagtc 533

Qy     984 cgtggtcaaccttgtggattctatttacaaggagcagctcaacaccagggttgtcctggt 1043
           |||||| ||||| |  |||   || ||||||||||||||||||||  |  | ||||||||
Db     534 cgtggtgaacctggccgatgtgatatacaaggagcagctcaacactcgcatcgtcctggt 593

Qy    1044 ggctgtagagacctggactgagaaggatcagattgacatcaccaccaaccctgtgcagat 1103
           ||  | || || ||| | ||   |||  ||||  |  |        |||   || |||
Db     594 tgccatggaaacatgggcagatggggacaagatccaggtgcaggatgacctcctggagac 653

Qy    1104 gctccatgagttctcaaaatac---cggcagcgcattaagcagcatgctgatgctgtgca 1160
           ||      | ||      |||   ||| || |  |     |||    | ||||   ||
Db     654 cctggcccggctcatggtctaccgacgggagggtctgcctgagcccagtaatgccaccca 713

Qy    1161 cctcatctcgcgggtgacatttcactataagagaagcagtctgagttacttggaggtgt 1220
           |||| |||||  |   ||| || ||    | ||| ||| |       ||| | || ||  |
Db     714 cctcttctcgggcaggaccttccagagcacgagcagcggggcagcctacgtgggggcat 773

Qy    1221 ctgttctcgcacaagaggagttggtgtgaatgagtatggtcttccaatggcagtggcaca 1280
           || || |   |    || |  |||||||| ||||| ||         |||  ||||
Db     774 atgctccctgtcccatggcgggggtgtgaacgagtacggcaacatgggggcgatggccgt 833

Qy    1281 agtattatcgcagagcctggctcaaaaccttggaatccaatgggaaccttctagcagaaa 1340
           |  | |||| |||| || ||||| || ||    ||| |        | ||
Db     834 gacccttgcccagacgctgggacagaacctgggcatgatgtggaacaaacaccggagctc 893

Qy    1341 gcca---aaatgtgactgcacagaatcctggggtggctgcatcatggaggaaacaggggt 1397
           | ||    | ||  | ||  ||||   ||||  ||||||||||||||||| || ||| |
Db     894 ggcaggggactgcaagtgtccagacatctggctgggctgcatcatggaggacactgggtt 953

Qy    1398 gtcccattctcgaaaattttcaaagtgcagcattttggagtatagagacttttttacagag 1457
           | ||   | || || || ||    ||||||||    ||||| |   | ||| | |||
Db     954 ctacctgccccgcaagttctctcgctgcagcatcgacgagtacaaccagtttctgcagga 1013

Qy    1458 aggaggtggagcctgccttttcaacaggccaacaaagctatttgagcccacggaatgtgg 1517
           || |||||   |||||| |||||||| |||    |||||  | || ||| | || || ||
Db    1014 gggtggtggcagctgcctcttcaacaagcccctcaagctcctggaccccccagagtgcgg 1073

Qy    1518 aaatggatacgtggaagctggggaggagtgtgattgtggttttcatgtggaatgc----- 1572
           || || | |||||| || |||||||||||| || || || |       ||| |||
Db    1074 gaacggcttcgtggaggcaggggaggagtgcgactgcggctcggtgcaggagtgcagccg 1133

Qy    1573 ----tatggattatgctgtaagaaatgttccctctccaacggggctcactgcagcgacgg 1628
               |||    ||||| |||||||| ||||  |  |||  ||    ||||||||||
Db    1134 cgcaggtggcaactgctgcaagaaatgcaccctgactcacgacgccatgtgcagcgacgg 1193

Qy    1629 gccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgccgggatgc 1688
           || ||||||| |          ||     |  | |||||  ||       |||||  || ||
Db    1194 gctctgctgtcgccgc------tgcaagtacgaaccacggggtgtgtcctgccgagaggc 1247

Qy    1689 tgtgaacgagtgtgatattactgaatattgtactggagactctggtcagtgcccaccaaa 1748
           ||||||||||| || ||  | ||    || || || |||||| | |||||||| || ||
Db    1248 cgtgaacgagtgcgacatcgcggagacctgcaccggggactctagccagtgcccgcctaa 1307

Qy    1749 tcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctacaatggcga 1808
           || || ||||  ||||| ||    ||  | ||  | ||||||||||||||   ||
Db    1308 cctgcacaagctggacggttactactgtgaccatgagcagggccgctgctacggaggtcg 1367

Qy    1809 gtgcaagaccagagacaaccagtgtcagtacatctggggaacaaaggctgcagggtctga 1868
           |||||  ||| | |||   ||||| |||    | |||||      |||||      ||||
Db    1368 ctgcaaaacccgggaccggcagtgccaggttctttggggccatgcggctg------ctga 1421

Qy    1869 caagttctgctatgaaaagctgaatacagaaggcactgagaagggaaactgcgggaagga 1928
           |||||||| || |||||||||   || || || |||   || | ||| |||    |
Db    1422 tcgcttctgctacgagaagctgaatgtggaggggacggagcgtgggagctgtgggcgcaa 1481

Qy    1929 tggagaccggtggattcagtgcagcaaacatgatgtgttctgtggattcttactctgtac 1988
           |||  | | ||| | |||||||| || || || ||| | ||||| ||| | ||||||
Db    1482 gggatccggctgggtccagtgcagtaagcaggacgtgctgtgtggcttcctcctctgtgt 1541

Qy    1989 caatcttactcgagctccacgtattggtcaacttcagggtgagatcattccaacttcctt 2048
           |||  |  || ||||||| ||  | ||  | ||   ||| || |||| |       ||||
Db    1542 caacatctctggagctcctcggctaggggacctggtgggagacatcagtagtgtcacctt 1601

Qy    2049 ctaccatcaaggccgggtgattgactgcagtggtgcccatgtagttttagatgatgatac 2108
           ||||||  || |||  || | | |||||||| || | ||| ||     | |  || |   |
Db    1602 ctaccaccagggcaaggagctggactgcaggggaggccacgtgcagctggcggacggctc 1661

Qy    2109 ggatgtgggctatgtagaagatggaacgccatgtggcccgtctatgatgtgtttagatcg 2168
           ||  || |||||||| || ||||| ||  | || || ||    ||| ||||  | || |
Db    1662 tgacctgagctatgtggaggatggcacagcctgcgggcctaacatgttgtgcctggacca 1721

Qy    2169 gaagtgcctacaaattcaagccctaaatatgagcagctgtccactcgattccaagggtaa 2228
           |||||  | |  |   ||| | ||  | |||| ||| ||   |  |    || |
Db    1722 tcgctgcctgccagcttctgccttcaacttcagcacctgccccggcagtggggagcgccg 1781

Qy    2229 agtctgttcgggccatggggtgtgtagtaatgaagccacctgcatttgtgatttcacctg 2288
           | || ||   ||| ||||| || || |||||||  |  ||||| ||| |      |||
Db    1782 gatttgctcccaccacggggtctgcagcaatgaagggaagtgcatctgtcagccagactg 1841

Qy    2289 ggcagggacagattgcagtatccgggatccagt---taggaaccttcacccccccaagga 2345
           | |||| | ||| ||||||||||   | ||  |    | |  ||  | | |     ||
Db    1842 gacaggcaaagactgcagtatccataacccccgcccacgtccccacccacgggggagac 1901

Qy    2346 tgaaggacccaagggtcctagtgccaccaatctcataataggctccatcgctggtgccat 2405
           ||  ||   || |||||  || | ||||||  |||| || ||||||||||||||  ||  |
Db    1902 ggagagatataaaggtcccagcggcaccaacatcatcattggctccatcgctggggctgt 1961

Qy    2406 cctggtagcagctattgtccttgggggcacaggctggggatttaaaaatgtcaagaagag 2465
           ||||||  |||||    |||||  || ||||| ||||||||||||||||  |       |
Db    1962 cctggttgcagcctacgtcctgggcggcacgggctggggatttaaaaacattcgccgagg 2021

Qy    2466 aaggttcg 2473
           |||||  ||
Db    2022 aaggtccg 2029
```

```
RESULT  10
AAQ76120
ID   AAQ76120 standard; cDNA to mRNA; 2923 BP.
XX
AC   AAQ76120;
XX
DT   01-AUG-1995  (first entry)
XX
DE   Human fetal cerebellum MDC cDNA.
XX
KW   MDC protein; breast cancer; mamma carcinoma; ovary cancer;
KW   chromosome-17; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   5'UTR           1..27
FT                   /*tag=  a
FT   CDS             28..1602
FT                   /*tag=  b
FT   3'UTR           1600..2923
FT                   /*tag=  c
XX
PN   EP633268-A.
XX
PD   11-JAN-1995.
XX
```

PF 13-MAY-1994; 94EP-0107487.
XX
PR 14-MAY-1993; 93JP-0136602.
PR 22-SEP-1993; 93JP-0257455.
PR 23-FEB-1994; 94JP-0049904.
PR 12-APR-1994; 94JP-0073328.
XX
PA (CANC-) CANCER INST.
PA (EISA) EISAI CO LTD.
XX
PI Emi M, Nakamura Y;
XX
DR WPI; 1995-038478/06.
DR P-PSDB; AAR67758.
XX
[redacted]Novel MDC protein and DNA encoding it - used to develop prods. for the study, diagnosis and therapy of cancers, partic. breast and ovarian cancer
XX
PS Disclosure; Page 86-89; 123pp; English.
XX
CC A detailed map of human chromosome-17 was constructed to analyze the
CC chromosome in breast and ovarian cancer tissues. 2 Novel cDNAs
CC (given in AAQ76120-21) were obtained from fetal cerebellum that
CC encoded novel MDC proteins (AAR67758-59, respectively). 5'-RACE and
CC RT-PCR revealed the sequence given in AAQ76122, encoding the MDC given
CC in AAR75352. Sequences common for these clones are given in AAQ76119
CC and AAR66757, respectively. A genomic DNA sequence (AAQ76124) for MDC
CC was obtained from cosmid clone cCI17-904.
XX
SQ Sequence 2923 BP; 607 A; 893 C; 887 G; 536 T; 0 other;

Query Match 20.5%; Score 513.2; DB 16; Length 2923;
Best Local Similarity 57.6%; Pred. No. 7.3e-130;
Matches 1116; Conservative 0; Mismatches 773; Indels 49; Gaps 9;

```
Qy    573 gggtggagagcactgttactaccatggaagcatcagaggcgtcaaagactccaaggtggc 632
Db    114 ggctggagaccactgctactaccaggggaagctccgggggaacccgcactccttcgccgc 173

Qy    633 tctgtcaacctgcaatggacttcatggcatgtttgaagatgataccttcgtgtatatgat 692
Db    174 cctctccacctgccaggggctgcatggggtcttctctgatgggaacttgacttacatcgt 233

      693 agagccactagagctggttcatgatgagaaaagcac------aggtcgaccacatataat 746
      234 ggagccccaagaggtggctggaccttggggagcccctcagggaccccttccccacctcat 293

Qy    747 ccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcactatggaaagagg 806
Db    294 ttaccggaccctctcctcccagatcccctcggatgcagggaaccaggctgcctgtttgc 353

Qy    807 tgaccagtggccctttctctctgaattacagtggttgaaaagaaggaagaagagcagtgaa 866
Db    354 tgtgcctgcccagtcggctcctccaaaccggccgaggctgagaaggaaaaggcaggtccg 413

Qy    867 tccatcacgtggtatatttga---agaaatgaaatatttggaacttatgattgttaatga 923
Db    414 ccggggccaccctacagtgcacagtgaaaccaagtatgtggagctaattggatcaacga 473

Qy    924 tcacaaaacgtataagaagcatcgctcttctcatgcacataccaacaacttttgcaaagtc 983
Db    474 ccaccagctgttcgagcagatgcgacagtcggtggtcctcaccagcaacttttgccaagtc 533

Qy    984 cgtggtcaaccttgtggattctatttacaaggagcagctcaacaccagggttgtcctggt 1043
Db    534 cgtggtgaacctggccgatgtgatatacaaggagcagctcaacactcgcatcgtcctggt 593

Qy   1044 ggctgtagagacctggactgagaaggatcagattgacatcaccaccaaccctgtgcagat 1103
Db    594 tgccatggaaacatgggcagatggggacaagatccaggtgcaggatgacctcctggagac 653

Qy   1104 gctccatgagttctcaaaatac---cggcagcgcattaagcagcatgctgatgctgtgca 1160
Db    654 cctggcccggctcatggtctaccgacgggagggtctgcctgagcccagtaatgccaccca 713

Qy   1161 cctcatctcgcgggtgacatttcactataagagaagcagtctgagttactttggaggtgt 1220
Db    714 cctcttctcgggcaggaccttccagagcacgagcagcggggcagcctacgtgggggcat 773

Qy   1221 ctgttctcgcacaagaggagttggtgtgaatgagtatggtcttccaatggcagtggcaca 1280
Db    774 atgctccctgtcccatggcgggggtgtgaacgagtacggcaacatgggggcgatggccgt 833

Qy   1281 agtattatcgcagagcctggctcaaaaccttggaatccaatgggaaccttctagcagaaa 1340
Db    834 gaccctgcccagacgctgggacagaacctgggcatgatgtggaacaaacaccggagctc 893

Qy   1341 gcca---aaatgtgactgcacagaatcctggggtggctgcatcatggaggaaacaggggt 1397
Db    894 ggcaggggactgcaagtgtccagacatctggctgggctgcatcatggaggacactgggtt 953

Qy   1398 gtcccattctcgaaaattttcaaagtgcagcattttggagtatagagacttttacagag 1457
Db    954 ctacctgccccgcaagttctctcgctgcagcatcgacgagtacaaccagtttctgcagga 1013

Qy   1458 aggaggtggagcctgccttttcaacaggccaacaaagctatttgagcccacggaatgtgg 1517
Db   1014 gggtggtggcagctgcctcttcaacaagcccctcaagctcctggacccccagagtgcgg 1073

Qy   1518 aaatggatacgtggaagctggggaggagtgtgattgtggttttcatgtggaatgc----- 1572
Db   1074 gaacggcttcgtggaggcaggggaggagtgcgactgcggctcggtgcaggagtgcagccg 1133

Qy   1573 ----tatggattatgctgtaagaaatgttccctctccaacggggctcactgcagcgacgg 1628
Db   1134 cgcaggtggcaactgctgcaagaaatgcaccctgactcacgacgccatgtgcagcgacgg 1193

Qy   1629 gccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgccgggatgc 1688
Db   1194 gctctgctgtcgccgc------tgcaagtacgaaccacggggtgtgtcctgccgagaggc 1247

Qy   1689 tgtgaacgagtgtgatattactgaatattgtactggagactctggtcagtgcccaccaaa 1748
Db   1248 cgtgaacgagtgcgacatcgcggagacctgcaccggggactctagccagtgcccgcctaa 1307

Qy   1749 tcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctacaatggcga 1808
Db   1308 cctgcacaagctggacggttactactgtgaccatgagcagggccgctgctacggaggtcg 1367

Qy   1809 gtgcaagaccagagacaaccagtgtcagtacatctggggaacaaaggctgcagggtctga 1868
Db   1368 ctgcaaaacccgggaccggcagtgccaggttctttggggccatgcggctg------ctga 1421

Qy   1869 caagttctgctatgaaaagctgaatacagaaggcactgagaagggaaactgcgggaagga 1928
Db   1422 tcgcttctgctacgagaagctgaatgtggaggggacggagcgtgggagctgtgggcgcaa 1481

Qy   1929 tggagaccggtggattcagtgcagcaaaca----------tgatgtgttctgtggattct 1978
Db   1482 gggatccggctgggtccagtgcagtaagcagccccaacagggacgtgctgtgtggcttcc 1541

Qy   1979 tactctgtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatcattc 2038
Db   1542 tcctctgtgtcaacatctctggagctcctcggctaggggacctggtgggagacatcagta 1601

Qy   2039 caacttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagttttag 2098
Db   1602 gtgtcaccttctaccaccagggcaaggagctggactgcaggggaggccacgtgcagctgg 1661

Qy   2099 atgatgatacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatgatgt 2158
Db   1662 cggacggctctgacctgagctatgtggaggatggcacagcctgcgggcctaacatgttgt 1721

Qy   2159 gtttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactcgatt 2218
```

```
Db   1722 gcctggaccatcgctgcctgccagcttctgccttcaacttcagcacctgccccggcagtg 1781

Qy   2219 ccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatttgtg 2278

Db   1782 gggagcgccggatttgctcccaccacggggtctgcagcaatgaagggaagtgcatctgtc 1841

Qy   2279 atttcacctgggcagggacagattgcagtatccgggatccagt---taggaaccttcacc 2335

Db   1842 agccagactggacaggcaaagactgcagtatccataacccctgcccacgtccccaccca 1901

Qy   2336 cccccaaggatgaaggacccaagggtcctagtgccaccaatctcataataggctccatcg 2395

Db   1902 cgggggagacggagagatataaaggtcccagcggcaccaacatcatcattggctccatcg 1961

Qy   2396 ctggtgccatcctggtagcagctattgtccttgggggcacaggctggggatttaaaaatg 2455

Db   1962 ctggggctgtcctggttgcagccatcgtcctgggcggcacgggctggggatttaaaaaca 2021

Qy   2456 tcaagaagagaaggttcg 2473

      022 ttcgccgaggaaggtccg 2039
```

RESULT 11
AAC04043
ID AAC04043 standard; cDNA; 446 BP.
XX
AC AAC04043;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 4041.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
DR P-PSDB; AAG04037.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 1; SEQ ID 4041; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors.
XX
SQ Sequence 446 BP; 149 A; 100 C; 91 G; 106 T; 0 other;

Query Match 14.8%; Score 369.2; DB 21; Length 446;
Best Local Similarity 96.5%; Pred. No. 5.8e-91;
Matches 388; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

```
Qy    281 aacagaatagcagcagtaatatcagttacagcaatgcaatgcagaaagaaatcacactgc 340

Db      1 aacagaatagcagcagtaatatcagttacagcaatgcaatgcagaaagaaatcacactgc 60

Qy    341 cttcaagactcatatattacatcaaccaagactcggaaagcccttatcacgttcttgaca 400

Db     61 cttcaagactcatatattacatcaaccaagactcggaaagcccttatcacgttcttgaca 120

Qy    401 caaaggcaagacaccagcaaaaacataataaggctgtccatctggcccaggcaagcttcc 460

Db    121 caaaggcaagacaccagcaaaaacataataaggctgtccatctggcccaggcaagcttcc 180

Qy    461 agattgaagccttcggctccaaattcattcttgacctcatactgaacaatggtttgttgt 520

Db    181 agattgaagccttcggctccaaattcattcttgacctcatactgaacaatggtttgttgt 240

Qy    521 cttctgattatgtggagattcactacgaaaatgggaaaccacagtactctaagggtggag 580

Db    241 cttctgattatgtggagattcactacgaaaatgggaaaccacagtactctaagggtggag 300

Qy    581 agcactgttactaccatggaagcatcagaggcgtcaaagactccaa-ggtggctctgtca 639

Db    301 agcactgttactaccatggaagcatcagaggcgtcaaagactccaagggtggctctgtca 360

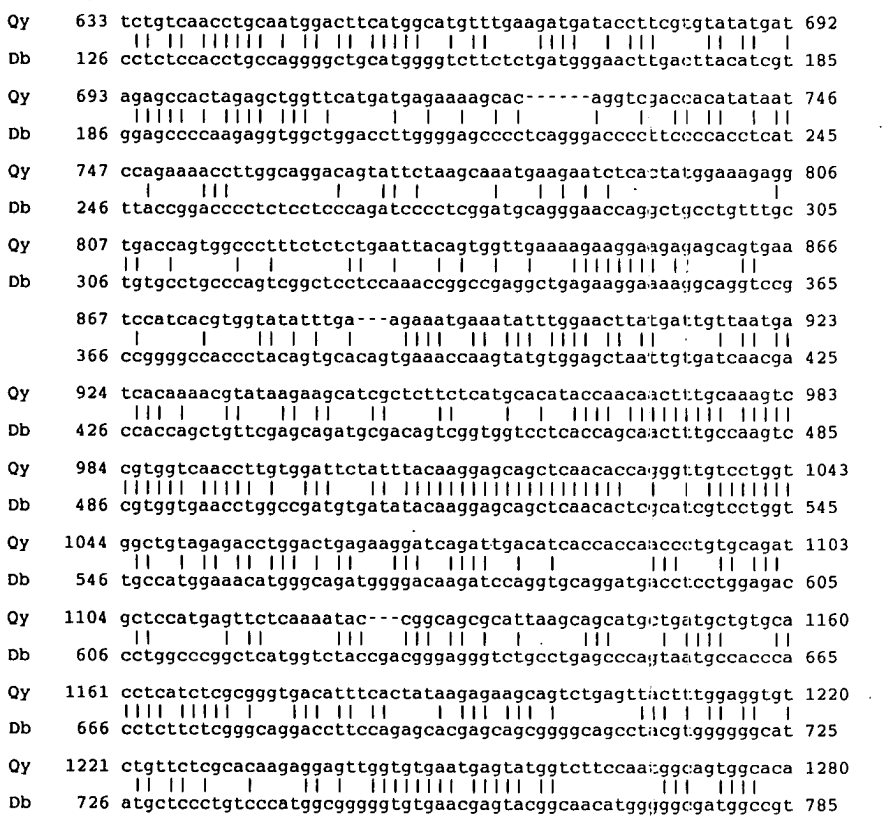
Qy    640 acctgcaatggacttcatggcatgtttgaagatgataccttc 681

Db    361 acctgcaatggacttcccgggatgggagaaagagagagactc 402
```

RESULT 12
AAQ76119
ID AAQ76119 standard; cDNA to mRNA; 1464 BP.
XX
AC AAQ76119;
XX
DT 01-AUG-1995 (first entry)
XX
DE Human fetal cerebellum MDC cDNA.
XX
KW MDC protein; breast cancer; mamma carcinoma; ovary cancer;
KW chromosome-17; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..1464
FT /*tag= a
XX
PN EP633268-A.
XX
PD 11-JAN-1995.
XX
PF 13-MAY-1994; 94EP-0107487.
XX
PR 14-MAY-1993; 93JP-0136602.
PR 22-SEP-1993; 93JP-0257455.
PR 23-FEB-1994; 94JP-0049904.
PR 12-APR-1994; 94JP-0073328.
XX
PA (CANC-) CANCER INST.
PA (EISA) EISAI CO LTD.
XX
PI Emi M, Nakamura Y;
XX

DR WPI; 1995-038478/06.
DR P-PSDB; AAR67757.
XX
PT Novel MDC protein and DNA encoding it - used to develop prods.
PT for the study, diagnosis and therapy of cancers, partic. breast
PT and ovarian cancer
XX
PS Disclosure; Page 82-85; 123pp; English.
XX
CC A detailed map of human chromosome-17 was constructed to analyze the
CC chromosome in breast and ovarian cancer tissues. 2 Novel cDNAs
CC (given in AAQ76120-21) were obtained from fetal cerebellum that
CC encoded novel MDC proteins (AAR67758-59, respectively). 5'-RACE and
CC RT-PCR revealed the sequence given in AAQ76122, encoding the MDC given
CC in AAR75352. Sequences common for these clones are given in AAQ76119
CC and AAR66757, respectively. A genomic DNA sequence (AAQ76124) for MDC
CC was obtained from cosmid clone cCI17-904.
XX
SQ Sequence 1464 BP; 294 A; 445 C; 473 G; 252 T; 0 other;

Query Match 13.5%; Score 338.4; DB 16; Length 1464;
Best Local Similarity 56.2%; Pred. No. 2.8e-82;
Matches 793; Conservative 0; Mismatches 581; Indels 36; Gaps 7;



RESULT 13
AAC85515
ID AAC85515 standard; cDNA; 2406 BP.
XX
AC AAC85515;
XX
DT 16-MAY-2001 (first entry)
XX
DE Degenerate first splice variant of MAPP cDNA, zdint2.
XX
KW Splice variant; zdint2; mammalian adhesion protease peptide; MAPP;
KW testis; ovary; prostate; small intestine; colon; stomach; thyroid;
KW spinal cord; lymph node; trachea; heart; wound healing; apoptosis;
KW neurogenesis; tumor proliferation; ischemia reperfusion; inflammation;
KW immunologic recognition; gamete maturation; platelet aggregation;
KW infarction; brain; cancer; Alzheimer's disease; multiple sclerosis;
KW congestive heart failure; ss.
XX
OS Homo sapiens.
XX
PN WO200109293-A2.
XX
PD 08-FEB-2001.
XX

PF 02-AUG-2000; 2000WO-US21085.
XX
PR 03-AUG-1999; 99US-0368070.
XX
PA (ZYMO) ZYMOGENETICS INC.
XX
PI Sheppard PO, Baindur N, Bishop PD;
XX
DR WPI; 2001-202662/20.
DR P-PSDB; AAB47105.
XX
PT Mammalian adhesion protease peptides useful for delivery of therapeutic
PT agents, for identifying agonists and antagonists and treating disorders
PT of brain, heart tissue and platelet aggregation -
XX
PS Disclosure; Page 99-100; 106pp; English.
XX
CC This sequence shows a degenerate version of a splice variant of the
CC zdint2 cDNA which encodes an isolated mammalian adhesion protease
CC peptide (MAPP). Analysis of tissue distribution of MAPP showed a
CC transcript of approx. 4.4kb with a strong signal in testes, ovary,
CC prostate, small intestine and colon, and a fainter signal in stomach,
CC thyroid, spinal cord, lymph node and trachea. Also there were two
CC transcripts, approx. 4kb and 4.4kb, both showing medium strength signal
CC in heart tissue. Peptide fragments of MAPP are useful for modulating
CC cell-cell interactions, where the cells are derived from tissues
CC showing a strong level of expression. MAPP proteins are useful for
CC delivery of therapeutic agents such as proteases, radionuclides,
CC chemotherapy agents and small molecules and for identifying cells,
CC tissues, cell lines which respond to a MAPP-stimulated pathway. MAPP
CC polypeptides are also useful for identifying agonists and antagonists,
CC to prepare antibodies and to study proliferation or differentiation in
CC e.g. testes, ovary, prostate, small intestine, colon and spinal cord.
CC The disintegrin loop of MAPP is useful in assays and treatment of
CC disorders of the spinal cord, heart, aorta, colon, bladder, small
CC intestine, uterus, prostate, stomach, testis, ovary, mammary gland,
CC appendix, lung, trachea, fetal lung and placenta. MAPP polypeptides,
CC antagonists and agonists are useful for promoting wound healing,
CC studying cell-cell interactions, apoptosis, neurogenesis, tumor
CC proliferation, repair and remodeling of ischemia reperfusion and
CC inflammation in vitro and in vivo, also useful in repair and remodeling
CC after an ischemic event, modulating immunologic recognition, gamete
CC maturation and/or platelet aggregation. Agonists are useful in
CC promoting the growth and development of cells of the myeloid and
CC lymphoid lineages in culture. Antibodies to MAPP are useful for tagging
CC cells that express MAPP, isolating MAPP by affinity purification,
CC determining circulating levels of MAPP polypeptides, for detecting or
CC quantitating soluble MAPP as marker of underlying pathology or disease
CC and as neutralizing antibodies or as antagonists to block MAPP in vitro
CC and in vivo. MAPP polypeptides, nucleic acids and/or antibodies are
CC useful in the treatment of disorders associated with infarct in brain
CC or heart tissue and/or platelet aggregation and as diagnostic
CC indicators of cancer. Disorders diagnosed or treated or prevented with
CC MAPP polypeptides, include Alzheimer's disease, tumor formation,
CC multiple sclerosis, congestive heart failure, ischemic reperfusion or
CC infarct and degenerative diseases.
XX
SQ Sequence 2406 BP; 281 A; 327 C; 511 G; 269 T; 1018 other;

Query Match 6.2%; Score 154.8; DB 22; Length 2406;
Best Local Similarity 30.7%; Pred. No. 4.9e-32;
Matches 417; Conservative 187; Mismatches 726; Indels 30; Gaps 7;

```
Qy    895 aaatatttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctcttct 954
          ||:||::| ||::|    ||:||    ||:||:|    | :  | :  | :     :
Db    628 aartayytngarytntayathgtngcngaycayacnytnttyytnacnmgncaymgnaay 687

Qy    955 catgcacataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttacaag 1014
          :    ||:|| ||  |             || |  ||:  :|| ||:    | :  :
Db    688 ytnaaycayacnaarcarmgnytnytngargtngcnaaytaygtngaycarytnytnm-- 745

Qy   1015 gagcagctcaacaccagggttgtcctggtggctgtagagacctggactgagaaggatcag 1074
                :|  |:| :  :|| |  :|    |   | ||:   ||||| ||::  ||::
Db    746 -gnacnytngayathcargtngcnytnacnggnytngargtntggacngarmgngaymgn 804

Qy   1075 attgacatcaccaccaaccctgtgcagatgctccatgagttctcaaaataccggcagcgc 1134
          :    | ||    |: |       |  :|    |  ||::   |:|  :| :   |
Db    805 wsnmgngtnacncargaygcnaaygcnacnytntgggcnttyytncartggmgnmgnggn 864

Qy   1135 at------taagcagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactat 1188
           |      |::  |    :||: : |  || :|  |  |  |     | ||:||
Db    865 ytntgggcncarmgnccncaygaywsngcncarytnytnacnggnmgngcnttycarggn 924

Qy   1189 aagagaagcagtctgagttactttggaggtgtctgttctcgcacaagaggagttggtgtg 1248
             |     | :|       | | :||  | ||:       :::  : |  || ||
Db    925 gcnacngtnggnytngcnccngtngarggnatgtgymgngcngarwsnwsnggnggngtn 984

Qy   1249 a------atgagtatggtcttccaatggcagtggcacaagtattatcgcagagcctggct 1302
          :      |: | :  |  :| || || |  |  ||        |  | ||     | |
Db    985 wsnacngaycaywsngarytnccnathggngcngcngcnacnatggcncaygarathggn 1044

Qy   1303 caaaaccttggaatccaatgggaaccttctagcagaaagccaaaatgtgactgcacagaa 1362
          || :  :| ||  |       || ||   : |  |       |:  |      | ||:
Db    1045 caywsnytnggnytnwsncaygayccngaygggntgytgygtngargcng---cngcngar 1101

Qy   1363 tcctggggtggctgcatcatggaggaaacaggggtgtcccattctcgaaaattttcaaag 1422
          ::  | ||  |:   ||    |  |  || ||     |   : | :|    ||:::
Db    1102 wsnggnggntgygtnatggcngcngcnacnggncayccnttyccnmgngtnttywsngcn 1161

Qy   1423 tgcagcattttggagtatagagacttttacagagaggaggtggagcctgccttttcaac 1482
          ||::: :      |::  :| |  ||:|| :  | :|| || || || ||::| :  ||:
Db    1162 tgywsnmgnmgncarytnmgngcnttyttymgnaarggnggnggngcntgyytnwsnaay 1221

Qy   1483 agg------ccaacaaagctatttgagcccacggaatgtggaaatggatacgtggaagct 1536
                ||      |   |    ||  |    ||:|| ||:|| | :|| ||:||
Db    1222 gcnccngayccnggnytnccngtnccnccngcnytntgyggnaayggnttygtngargcn 1281

Qy   1537 ggggaggagtgtgattgtggttttcatgtggaatgctatggattatgctgtaagaaatgt 1596
          || ||:||:||:||:||:||        :||:||:   |  :| ||:||:       :
Db    1282 ggngargartgygaytgyggncnggncargartgymgngayytntgytgyttygcncay 1341

Qy   1597 tccctctccaacggggctcactgcagcgacgggccctgctgtaacaatacctcatgtctt 1656
             :  :::    |  |           :|  | : |   :  :  :      ||::|
Db    1342 aaytgywsnytnmgnccnggngcncartgygcncayggngaytgytgygtnmgntgyytn 1401

Qy   1657 tttcagccacgagggtatgaatgccgggatgctgtgaacgagtgtgatattactgaatat 1716
          :|  |:||   ||      ||::|  | ||  ||   || ||:||: |  | ||:| :
Db    1402 ytnaarccngcnggngcnytntgymgncargcnatgggngaytgygayytnccngartty 1461

Qy   1717 tgtactggagactctggtcagtgcccaccaaatcttcataagcaagacggatatgcatgc 1776
          ||:|| ||    ::  : || ||:|| ||  |: |  |:   :  ||:|| :   | ||:
Db    1462 tgyacnggnacnwsnwsncaytgyccnccngaygtntayytnytngayggnwsnccntgy 1521

Qy   1777 aatcaaaatcagggccgctgctacaatggcgagtgcaagaccagagacaaccagtgtcag 1836
             :        ||   :||:|   |:|| |  ||:   ||    ||  | ||:||:||:
Db    1522 gcnmgnggnwsnggntaytgytgggayggngcntgyccnacnytngarcarcartgycar 1581

Qy   1837 tacatctggggaacaaaggctgcagggtctgacaagttctgctatgaaaagctgaataca 1896
          |  | |||||  |     :       |      |:   ||:| : |:    | ||:::
Db    1582 carytntggggnccnggnwsncayccngcnccngargcntgyttycargtngtnaaywsn 1641

Qy   1897 gaaggcactgagaagggaaactgcgggaaggatggaga---ccggtggattcagtgcagc 1953
          |  ||   :|   | || ||:||:||  |:||: : ||    |  |   | |  ||:
Db    1642 gcnggngaygcncayggnaaytgyggncargaywsngargggncayttyytnccntgygcn 1701

Qy   1954 aaacatgatgtgttctgtggattcttactctgtaccaatcttactcgagctccacgtatt 2013
          :  ||:|  :| ||:||    :| |  ||:         |  |   : :  :
Db    1702 ggnmgngaygcnytntgyggnaarytncartgycarggnggnaarccnwsnytnytngcn 1761

Qy   2014 ggtca---acttcagggtgagatcattccaacttccttctaccatcaaggccgggtgatt 2070
             ||     | |  |  || :  |       ::    :   ||  |:|       :
Db    1762 ccncayatggtnccngtngaywsnacngtncayytngaygggncargargtnacntgymgn 1821

Qy   2071 gactgcagtggtgcccatgtagttttagatgatgatacggatgtgggctatgtagaagat 2130
          |        |   |    |   :  ||:      |   | || :  || ||:
```

```
Db   1822 ggngcnytngcnytnccnwsngcncarytngayytnytnggnytnggnytngtngarccn 1881

Qy   2131 ggaacgccatgtggcccgtctatgatgtgtttagatcggaagtgcctacaaattcaagcc 2190
          || || | :||:|| ||    ||| | ||:  :   :| :  ||::  |:| :    :
Db   1882 ggnacncartgyggnccnmgnatggtntgycarwsnmgnmgntgymgnaaraaygcntty 1941

Qy   2191 ctaaatatgagcagctgtccactcgattccaagggtaaag 2230
          | : |  |    :| ||::    |  | : |  :  | |
Db   1942 cargarytncarmgntgyytnacngcntgycaywsncayg 1981


RESULT 14
AAC85516
ID   AAC85516 standard; cDNA; 2439 BP.
XX
AC   AAC85516;
XX
DT   16-MAY-2001  (first entry)
XX
DE   Degenerate second splice variant of MAPP cDNA, zdint2.
XX
KW   Splice variant; zdint2; mammalian adhesion protease peptide; MAPP;
KW   testis; ovary; prostate; small intestine; colon; stomach; thyroid;
KW   spinal cord; lymph node; trachea; heart; wound healing; apoptosis;
KW   neurogenesis; tumor proliferation; ischemia reperfusion; inflammation;
KW   immunologic recognition; gamete maturation; platelet aggregation;
KW   infarction; brain; cancer; Alzheimer's disease; multiple sclerosis;
KW   congestive heart failure; ss.
XX
OS   Homo sapiens.
XX
PN   WO200109293-A2.
XX
PD   08-FEB-2001.
XX
PF   02-AUG-2000; 2000WO-US21085.
XX
PR   03-AUG-1999;   99US-0368070.
XX
PA   (ZYMO ) ZYMOGENETICS INC.
XX
PI   Sheppard PO,  Baindur N,  Bishop PD;
XX
DR   WPI; 2001-202662/20.
DR   P-PSDB; AAB47106.
XX
PT   Mammalian adhesion protease peptides useful for delivery of therapeutic
PT   agents, for identifying agonists and antagonists and treating disorders
PT   of brain, heart tissue and platelet aggregation -
XX
PS   Disclosure; Page 100-01; 106pp; English.
XX
CC   This sequence shows a degenerate version of a splice variant of the
CC   zdint2 cDNA which encodes an isolated mammalian adhesion protease
CC   peptide (MAPP). Analysis of tissue distribution of MAPP showed a
CC   transcript of approx. 4.4kb with a strong signal in testes, ovary,
CC   prostate, small intestine and colon, and a fainter signal in stomach,
CC   thyroid, spinal cord, lymph node and trachea. Also there were two
CC   transcripts, approx. 4kb and 4.4kb, both showing medium strength signal
CC   in heart tissue.  Peptide fragments of MAPP are useful for modulating
CC   cell-cell interactions, where the cells are derived from tissues
CC   showing a strong level of expression.  MAPP proteins are useful for
CC   delivery of therapeutic agents such as proteases, radionuclides,
CC   chemotherapy agents and small molecules and for identifying cells,
CC   tissues, cell lines which respond to a MAPP-stimulated pathway. MAPP
CC   polypeptides are also useful for identifying agonists and antagonists,
CC   to prepare antibodies and to study proliferation or differentiation in
CC   e.g. testes, ovary, prostate, small intestine, colon and spinal cord.
CC   The disintegrin loop of MAPP is useful in assays and treatment of
CC   disorders of the spinal cord, heart, aorta, colon, bladder, small
CC   intestine, uterus, prostate, stomach, testis, ovary, mammary gland,
CC   appendix, lung, trachea, fetal lung and placenta. MAPP polypeptides,
CC   antagonists and agonists are useful for promoting wound healing,
CC   studying cell-cell interactions, apoptosis, neurogenesis, tumor
CC   proliferation, repair and remodeling of ischemia reperfusion and
CC   inflammation in vitro and in vivo, also useful in repair and remodeling
CC   after an ischemic event, modulating immunologic recognition, gamete
CC   maturation and/or platelet aggregation. Agonists are useful in
CC   promoting the growth and development of cells of the myeloid and
CC   lymphoid lineages in culture. Antibodies to MAPP are useful for tagging
CC   cells that express MAPP, isolating MAPP by affinity purification,
CC   determining circulating levels of MAPP polypeptides, for detecting or
CC   quantitating soluble MAPP as marker of underlying pathology or disease
CC   and as neutralizing antibodies or as antagonists to block MAPP in vitro
CC   and in vivo. MAPP polypeptides, nucleic acids and/or antibodies are
CC   useful in the treatment of disorders associated with infarct in brain
CC   or heart tissue and/or platelet aggregation and as diagnostic
CC   indicators of cancer. Disorders diagnosed or treated or prevented with
CC   MAPP polypeptides, include Alzheimer's disease, tumor formation,
CC   multiple sclerosis, congestive heart failure, ischemic reperfusion or
CC   infarct and degenerative diseases.
XX
SQ   Sequence 2439 BP; 291 A; 327 C; 519 G; 278 T; 1024 other;


  Query Match             6.2%;  Score 154.8;  DB 22;  Length 2439;
  Best Local Similarity   30.7%;  Pred. No. 4.9e-32;
  Matches  417;  Conservative  187;  Mismatches  726;  Indels   30;  Gaps    7;

Qy    895 aaatatttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctcttct 954
          ||:||::| ||::|    ||:||    ||:||:|     | :   |  :  | :     :
Db    628 aartayytngarytntayathgtngcngaycayacnytnttyytnacnmgncaymgnaay 687

Qy    955 catgcacataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttacaag 1014
          :     ||:|| ||  |              || |  ||:  :|| ||:    | :  :
Db    688 ytnaaycayacnaarcarmgnytnytngargtngcnaaytaygtngaycarytnytnm-- 745

Qy   1015 gagcagctcaacaccagggttgtcctggtggctgtagagacctggactgagaaggatcag 1074
               :|  |:| :  :|| |  :|    |   | ||:   ||||| ||::  ||::
Db    746 -gnacnytngayathcargtngcnytnacnggnytngargtntggacngarmgngaymgn 804

Qy   1075 attgacatcaccaccaaccctgtgcagatgctccatgagttctcaaaataccggcagcgc 1134
          :    | ||     |: |      |  :|    |  ||:;   |:|  :| :   |
Db    805 wsnmgngtnacncargaygcnaaygcnacnytntgggcnttyytncartggmgnmgnggn 864

Qy   1135 at------taagcagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactat 1188
           |       |::  |    :||: : |  || :|  |  |  |     | ||:||
Db    865 ytntgggcncarmgncncaygaywsngcncarytnytnacnggnmgngcnttycarggn 924

Qy   1189 aagagaagcagtctgagttactttggaggtgtctgttctcgcacaagaggagttggtgtg 1248
             |     | :|       | | :||  | ||:        :::  : |  || ||
Db    925 gcnacngtnggnytngcnccngtngarggnatgtgymgngcngarwsnwsnggnggngtn 984

Qy   1249 a------atgagtatggtcttccaatggcagtggcacaagtattatcgcagagcctggct 1302
          :      |: | :  |  :| || || |  |  ||       |  | ||     | |
Db    985 wsnacngaycaywsngarytnccnathggngcngcngcnacnatggcncaygarathggn 1044

Qy   1303 caaaaccttggaatccaatgggaaccttctagcagaaagccaaaatgtgactgcacagaa 1362
          || :  :| ||  |       || ||   : |  |        |:  |     | ||:
Db    1045 caywsnytnggnytnwsncaygayccngayggngtgytgygtngargcng---cngcngar 1101

Qy   1363 tcctggggtggctgcatcatggaggaaacaggggtgtcccattctcgaaaattttcaaag 1422
          ::  | ||  |:   ||    |  |  || ||     |   : | :|    ||:::
Db    1102 wsnggnggntgygtnatggcngcngcnacnggncayccnttyccnmgngtnttywsngcn 1161

Qy   1423 tgcagcattttggagtatagagacttttacagagaggaggtggagcctgccttttcaac 1482
          ||::: :      |::  :| |  ||:|| :  | :|| || || || ||::| :  ||:
Db    1162 tgywsnmgnmgncarytnmgngcnttyttymgnaarggnggnggngcntgyytnwsnaay 1221

Qy   1483 agg------ccaacaaagctatttgagcccacggaatgtggaaatggatacgtggaagct 1536
                   ||       |   |    ||  |    ||:|| ||:|| | :|| ||:||
Db    1222 gcnccngayccnggnytnccngtnccnccngcnytntgyggnaayggnttygtngargcn 1281

Qy   1537 ggggaggagtgtgattgtggttttcatgtggaatgctatggattatgctgtaagaaatgt 1596
          || ||:||:||:||:||:||        :||:||:   |  :| ||:||:       :
Db    1282 ggngargartgygaytgyggnccnggncargartgymgngayytntgytgyttygcncay 1341
```

```
Qy   1597 tccctctccaacggggctcactgcagcgacgggccctgctgtaacaatacctcatgtctt 1656
Db   1342 aaytgywsnytnmgnccnggngcncartgygcncayggngaytgytgygtnmgntgyytn 1401

Qy   1657 tttcagccacgagggtatgaatgccgggatgctgtgaacgagtgtgatattactgaatat 1716
Db   1402 ytnaarccngcnggngcnytntgymgncargcnatgggngaytgygayytnccngartty 1461

Qy   1717 tgtactggagactctggtcagtgcccaccaaatcttcataagcaagacggatatgcatgc 1776
Db   1462 tgyacnggnacnwsnwsncaytgyccnccngaygtntayytnytngayggnwsnccntgy 1521

Qy   1777 aatcaaaatcagggccgctgctacaatggcgagtgcaagaccagagacaaccagtgtcag 1836
Db   1522 gcnmgnggnwsnggntaytgytgggayggngcntgyccnacnytngarcarcartgycar 1581

Qy   1837 tacatctggggaacaaaggctgcagggtctgacaagttctgctatgaaaagctgaatataca 1896
Db   1582 carytntggggnccnggnwsncayccngcnccngargcntgyttycargtngtnaaywsn 1641

     1897 gaaggcactgagaagggaaactgcgggaaggatggaga---ccggtggattcagtgcagc 1953
     1642 gcnggngaygcncayggnaaytgyggncargaywsngarggncayttyytnccntgygcn 1701

Qy   1954 aaacatgatgtgttctgtggattcttactctgtaccaatcttactcgagctccacgtatt 2013
Db   1702 ggnmgngaygcnytntgyggnaarytncartgycarggnggnaarccnwsnytnytngcn 1761

Qy   2014 ggtca---acttcagggtgagatcattccaacttccttctaccatcaaggccgggtgatt 2070
Db   1762 ccncayatggtnccngtngaywsnacngtncayytngayggncargargtnacntgymgn 1821

Qy   2071 gactgcagtggtgcccatgtagttttagatgatgatacggatgtgggctatgtagaagat 2130
Db   1822 ggngcnytngcnytnccnwsngcncarytngayytnytnggnytnggnytngtngarccn 1881

Qy   2131 ggaacgccatgtggcccgtctatgatgtgtttagatcggaagtgcctacaaattcaagcc 2190
Db   1882 ggnacncartgyggncnmgnatggtntgycarwsnmgnmgntgymgnaaraaygcntty 1941

Qy   2191 ctaaatatgagcagctgtccactcgattccaagggtaaag 2230
Db   1942 cargarytncarmgntgyytnacngcntgycaywsncayg 1981
```

RESULT 15
AAT86011
ID AAT86011 standard; cDNA; 2848 BP.
XX
AC AAT86011;
XX
[redacted]08-MAY-1998 (first entry)
XX
Human alpha meltrin cDNA fragment 2.
XX
KW Meltrin; myotube formation; muscle tissue; therapy; membrane protein;
KW myoblasts; adhesion; fusion; medicinal; muscle disorder; human; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..2061
FT /*tag= a
FT /product= alpha meltrin
FT /note= "partial coding sequence"
XX
PN WO9731109-A1.
XX
PD 28-AUG-1997.
XX
PF 17-OCT-1996; 96WO-JP03017.
XX
PR 23-FEB-1996; 96JP-0061756.
XX
PA (MOCH) MOCHIDA PHARM CO LTD.
XX
PI Fujisawa A, Mizushima C, Ogawa N, Shirakawa K, Yamakawa T;
XX
DR WPI; 1997-435161/40.
DR P-PSDB; AAW25719.
XX
PT Meltrin proteins involved with myotube formation in muscle - contain
PT metallo:proteinase and disintegrin domains
XX
PS Claim 5; Fig 15; 262pp; Japanese.
XX
CC AAT86010-T86012 encode fragments of a new human membrane protein,
CC alpha meltrin, which participates in the adhesion and fusion of
CC myoblasts in the course of myotube formation in muscle tissue. Meltrins
CC and their partial peptides can be used in the study of myotube formation,
CC and in medicinal compositions for muscle disorders. Meltrins contain an
CC intracellular domain, a membrane-penetrating domain, a fusion peptide
CC domain, a disintegrin domain, a metalloproteinase domain, a precursor
CC domain and a cysteine rich domain although these are not shown in the
CC specification. Meltrin antagonists can be used as bone absorption
CC inhibitors, e.g. for the treatment of osteoporosis and hypercalcaemia
CC and for prevention of cancer cell infiltration.
XX
SQ Sequence 2848 BP; 760 A; 706 C; 790 G; 592 T; 0 other;

Query Match 4.6%; Score 113.8; DB 18; Length 2848;
Best Local Similarity 47.4%; Pred. No. 8.4e-21;
Matches 479; Conservative 0; Mismatches 517; Indels 15; Gaps 4;

```
Qy   1235 gaggagttggtgtgaatgagtatggtcttccaatggcagtggcacaagtattatcgcaga 1294
Db    836 ggggaattgtcatggaccattcagacaatccccttggtgcagccgtgaccctggcacatg 895

Qy   1295 gcctggctcaaaaccttggaatccaatgggaaccttctagcagaaagccaaaatgtgact 1354
Db    896 agctgggccacaatttcgggatgaatcatgacacactggacaggggctgtagctgtcaaa 955

Qy   1355 gcacagaatcctggggtggctgcatcatggaggaaacaggggtgtcccattctcgaa--- 1411
Db    956 tggcggttgagaaaggaggctgcatcatgaacgcttccaccgggtacccatttcccatgg 1015

Qy   1412 aattttcaaagtgcagcattttggagtatagagacttttttacagagaggaggtggagcct 1471
Db   1016 tgttcagcagttgcagcaggaaggacttggagaccagcctggagaaaggaatgggggtgt 1075

Qy   1472 gccttttcaacaggccaacaa------agctatttgagcccacggaatgtggaaatggat 1525
Db   1076 gcctgtttaacctgccggaagtcagggagtctttcgggggccagaagtgtgggaacagat 1135

Qy   1526 acgtggaagctggggaggagtgtgattgtggttttcatgtggaatgctatggattatgct 1585
Db   1136 ttgtggaagaaggagaggagtgtgactgtggggagccagaggaatgtatgaatcgctgct 1195

Qy   1586 gtaagaaatgttccctctccaacggggctcactgcagcgacgggccctgctgtaacaata 1645
Db   1196 gcaatgccaccacctgtaccctgaagccggacgctgtgtgcgcacatgggctgtgctgtg 1255

Qy   1646 cctcatgtctttttcagccacgagggtatgaatgccgggatgctgtgaacgagtgtgata 1705
Db   1256 aagactgccagctgaagcctgcaggaacagcgtgcagggactccagcaactcctgtgacc 1315

Qy   1706 ttactgaatattgtactggagactctggtcagtgcccaccaaatcttcataagcaagacg 1765
Db   1316 tcccagagttctgcacaggggccagccctcactgcccagccaacgtgtacctgcacgatg 1375

Qy   1766 gatatgcatgcaatcaaaatcagggccgctgctacaatggcgagtgcaagaccagagaca 1825
Db   1376 ggcactcatgtcaggatgtggacggctactgctacaatggcatctgccagactcacgagc 1435
```

```
Qy   1826 accagtgtcagtacatctggggaacaaaggctgcagggtctgacaagttctgctatgaaa 1885
          | ||||||       ||||||| ||   |||  |    |      | |||||| ||| |
Db   1436 agcagtgtgtcacgctctggggaccaggtgctaaacctgcccctgggatctgctttgaga 1495

Qy   1886 agctgaatacagaaggcactgagaagggaaactgcgggaaggatgggagccggtggatt- 1944
            | ||| | | |||   |    | || ||||| || || |      |  | |   ||
Db   1496 gagtcaattctgcaggtgatccttatggcaactgtggcaaagtctcgaacagttcctttg 1555

Qy   1945 --cagtgcagcaaacatgatgtgttctgtggattcttactctgtaccaatcttactcgag 2002
             | |||   |     ||||    ||||||    | |  |||       | |  |
Db   1556 ccaaatgcgagatgagagatgctaaatgtggaaaaatccagtgtcaaggaggtgccagcc 1615

Qy   2003 ctccacgtattggtcaacttcagggtgagatcattccaacttccttctaccatcaaggcc 2062
            |||   ||||||     |   | |   ||     |||    |      || |||||
     1616 ggccagtcattggtaccaatgccgtttccatagaaacaaacatccccctgcagcaaggag 1675

     2063 gggtgattgactgcagtggtgcccatgtagttttagatgatgata---cggatgtgggct 2119
          |   ||||   ||| | ||  |||| ||    || |  ||||| |   ||||    ||
Db   1676 gccggattctgtgccgggggacccacgtgtacttgggcgatgacatgccggacccagggc 1735

Qy   2120 atgtagaagatggaacgccatgtggcccgtctatgatgtgtttagatcggaagtgcctac 2179
           |||    |  || ||    ||||         |  || ||  |  ||||  | || | |
Db   1736 ttgtgcttgcaggcacaaagtgtgcagatggaaaaatctgcctgaatcgtcaatgtcaaa 1795

Qy   2180 aaattcaagccctaaatatgagcagctgtccactcgattccaagggtaaag 2230
          | |||   | | |     |   |   ||| || |  | | | | || | ||
Db   1796 atattagtgtctttggggttcacgagtgtgcaatgcagtgccacggcagag 1846
```

RESULT 16
AAZ33953
ID AAZ33953 standard; cDNA; 3305 BP.
XX
AC AAZ33953;
XX
DT 07-DEC-1999 (first entry)
XX
DE Human PRO545 nucleotide sequence.
XX
KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein; ss.
XX
OS Homo sapiens.
XX
WO9946281-A2.

16-SEP-1999.

PF 08-MAR-1999; 99WO-US05028.
XX
PR 10-MAR-1998; 98US-0077450.
PR 11-MAR-1998; 98US-0077632.
PR 11-MAR-1998; 98US-0077641.
PR 11-MAR-1998; 98US-0077649.
PR 12-MAR-1998; 98US-0077791.
PR 13-MAR-1998; 98US-0078004.
PR 17-MAR-1998; 98US-0040220.
PR 20-MAR-1998; 98US-0078886.
PR 20-MAR-1998; 98US-0078910.
PR 20-MAR-1998; 98US-0078936.
PR 20-MAR-1998; 98US-0078939.
PR 25-MAR-1998; 98US-0079294.
PR 26-MAR-1998; 98US-0079656.
PR 27-MAR-1998; 98US-0079663.
PR 27-MAR-1998; 98US-0079664.
PR 27-MAR-1998; 98US-0079689.
PR 27-MAR-1998; 98US-0079728.
PR 27-MAR-1998; 98US-0079786.
PR 30-MAR-1998; 98US-0079920.
PR 30-MAR-1998; 98US-0079923.
PR 31-MAR-1998; 98US-0080105.
PR 31-MAR-1998; 98US-0080107.
PR 31-MAR-1998; 98US-0080165.
PR 31-MAR-1998; 98US-0080194.
PR 01-APR-1998; 98US-0080327.
PR 01-APR-1998; 98US-0080328.
PR 01-APR-1998; 98US-0080333.
PR 01-APR-1998; 98US-0080334.
PR 08-APR-1998; 98US-0081049.
PR 08-APR-1998; 98US-0081070.
PR 08-APR-1998; 98US-0081071.
PR 09-APR-1998; 98US-0081195.
PR 09-APR-1998; 98US-0081203.
PR 09-APR-1998; 98US-0081229.
PR 15-APR-1998; 98US-0081817.
PR 15-APR-1998; 98US-0081838.
PR 15-APR-1998; 98US-0081952.
PR 15-APR-1998; 98US-0081955.
PR 21-APR-1998; 98US-0082568.
PR 21-APR-1998; 98US-0082569.
PR 22-APR-1998; 98US-0082700.
PR 22-APR-1998; 98US-0082704.
PR 22-APR-1998; 98US-0082804.
PR 23-APR-1998; 98US-0082767.
PR 23-APR-1998; 98US-0082796.
PR 27-APR-1998; 98US-0083336.
PR 28-APR-1998; 98US-0083322.
PR 29-APR-1998; 98US-0083392.
PR 29-APR-1998; 98US-0083495.
PR 29-APR-1998; 98US-0083496.
PR 29-APR-1998; 98US-0083499.
PR 29-APR-1998; 98US-0083500.
PR 29-APR-1998; 98US-0083545.
PR 29-APR-1998; 98US-0083554.
PR 29-APR-1998; 98US-0083558.
PR 29-APR-1998; 98US-0083559.
PR 30-APR-1998; 98US-0083742.
PR 05-MAY-1998; 98US-0084366.
PR 06-MAY-1998; 98US-0084414.
PR 06-MAY-1998; 98US-0084441.
PR 07-MAY-1998; 98US-0084598.
PR 07-MAY-1998; 98US-0084600.
PR 07-MAY-1998; 98US-0084627.
PR 07-MAY-1998; 98US-0084637.
PR 07-MAY-1998; 98US-0084639.
PR 07-MAY-1998; 98US-0084640.
PR 07-MAY-1998; 98US-0084643.
PR 13-MAY-1998; 98US-0085323.
PR 13-MAY-1998; 98US-0085338.
PR 13-MAY-1998; 98US-0085339.
PR 15-MAY-1998; 98US-0085573.
PR 15-MAY-1998; 98US-0085579.
PR 15-MAY-1998; 98US-0085580.
PR 15-MAY-1998; 98US-0085582.
PR 15-MAY-1998; 98US-0085689.
PR 15-MAY-1998; 98US-0085697.
PR 15-MAY-1998; 98US-0085700.
PR 15-MAY-1998; 98US-0085704.
PR 18-MAY-1998; 98US-0086023.
PR 22-MAY-1998; 98US-0086392.
PR 22-MAY-1998; 98US-0086414.
PR 22-MAY-1998; 98US-0086430.
PR 22-MAY-1998; 98US-0086486.
PR 28-MAY-1998; 98US-0087098.
PR 28-MAY-1998; 98US-0087106.
PR 28-MAY-1998; 98US-0087208.
PR 30-JUL-1998; 98US-0094651.
PR 11-SEP-1998; 98US-0100038.
XX
PA (GETH) GENENTECH INC.
XX
PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
XX
DR WPI; 1999-551358/46.
DR P-PSDB; AAY41695.

```
XX
PT   New secreted and transmembrane polypeptides and their polynucleotides,
PT   useful for treating blood coagulation disorders, cancers and cellular
PT   adhesion disorders -
XX
PS   Claim 2; Fig 29; 530pp; English.
XX
CC   The present invention describes secreted and transmembrane polypeptides
CC   and their polynucleotides. The nucleotide sequences are useful as
CC   sources of probes, primers, for chromosome mapping, and for generation
CC   of antisense sequences. They can also be used to create transgenic
CC   animals. The proteins can be used to treat a variety of diseases and
CC   disorders, depending on their function. Diseases that may be treated
CC   include blood coagulation disorders, cancers and cellular adhesion
CC   disorders. They may also be used to raise antibodies. AAZ33891 to
CC   AAZ34338, and AAY41685 to AAY41774 represent polynucleotide and
CC   polypeptide sequence given in the exemplification of the present
CC   invention.
XX
SQ   Sequence 3305 BP; 836 A; 855 C; 945 G; 669 T; 0 other;


  ry Match              4.6%;  Score 113.8;  DB 20;  Length 3305;
  st Local Similarity   47.4%;  Pred. No. 9.1e-21;
  Matches  479;  Conservative    0;  Mismatches  517;  Indels   15;  Gaps    4;

Qy   1235 gaggagttggtgtgaatgagtatggtcttccaatggcagtggcacaagtattatcgcaga 1294
Db   1293 ggggaattgtcatggaccattcagacaatccccttggtgcagccgtgaccctggcacatg 1352

Qy   1295 gcctggctcaaaaccttggaatccaatgggaaccttctagcagaaagccaaaatgtgact 1354
Db   1353 agctgggccacaatttcgggatgaatcatgacacactggacaggggctgtagctgtcaaa 1412

Qy   1355 gcacagaatcctggggtggctgcatcatggaggaaacaggggtgtcccattctcgaa--- 1411
Db   1413 tggcggttgagaaaggaggctgcatcatgaacgcttccaccgggtacccatttcccatgg 1472

Qy   1412 aattttcaaagtgcagcattttggagtatagagacttttttacagagaggaggtggagcct 1471
Db   1473 tgttcagcagttgcagcaggaaggacttggagaccagcctggagaaaggaatgggggtgt 1532

Qy   1472 gccttttcaacaggccaacaa------agctatttgagcccacggaatgtggaaatggat 1525
Db   1533 gcctgtttaacctgccggaagtcagggagtctttcgggggccagaagtgtgggaacagat 1592

Qy   1526 acgtggaagctggggaggagtgtgattgtggttttcatgtggaatgctatggattatgct 1585
Db   1593 ttgtggaagaaggagaggagtgtgactgtggggagccagaggaatgtatgaatcgctgct 1652

Qy   1586 gtaagaaatgttccctctccaacggggctcactgcagcgacgggccctgctgtaacaata 1645
Db   1653 gcaatgccaccacctgtaccctgaagccggacgctgtgtgcgcacatgggctgtgctgtg 1712

Qy   1646 cctcatgtctttttcagccacgagggtatgaatgccgggatgctgtgaacgagtgtgata 1705
Db   1713 aagactgccagctgaagcctgcaggaacagcgtgcagggactccagcaactcctgtgacc 1772

Qy   1706 ttactgaatattgtactggagactctggtcagtgcccaccaaatcttcataagcaagacg 1765
Db   1773 tcccagagttctgcacaggggccagccctcactgcccagccaatgtgtacctgcacgatg 1832

Qy   1766 gatatgcatgcaatcaaaatcagggccgctgctacaatggcgagtgcaagaccagagaca 1825
Db   1833 ggcactcatgtcaggatgtggacggctactgctacaatggcatctgccagactcacgagc 1892

Qy   1826 accagtgtcagtacatctggggaacaaaggctgcagggtctgacaagttctgctatgaaa 1885
Db   1893 agcagtgtgtcacgctctggggaccaggtgctaaacctgcccctgggatctgctttgaga 1952

Qy   1886 agctgaatacagaaggcactgagaagggaaactgcgggaaggatggagaccggtggatt- 1944
Db   1953 gagtcaattctgcaggtgatccttatggcaactgtggcaaagtctcgaagagttcctttg 2012

Qy   1945 --cagtgcagcaaacatgatgtgttctgtggattcttactctgtaccaatcttactcgag 2002
Db   2013 ccaaatgcgagatgagagatgctaaatgtggaaaaatccagtgtcaaggaggtgccagcc 2072

Qy   2003 ctccacgtattggtcaacttcagggtgagatcattccaacttccttctaccatcaaggcc 2062
Db   2073 ggccagtcattggtaccaatgccgtttccatagaaacaaacatccctctgcagcaaggag 2132

Qy   2063 gggtgattgactgcagtggtgcccatgtagttttagatgatgata---cggatgtgggct 2119
Db   2133 gccggattctgtgccgggggacccacgtgtacttgggcgatgacatgccggacccagggc 2192

Qy   2120 atgtagaagatggaacgccatgtggcccgtctatgatgtgtttagatcggaagtgcctac 2179
Db   2193 ttgtgcttgcaggcacaaagtgtgcagatggaaaaatctgcctgaatcgtcaatgtcaaa 2252

Qy   2180 aaattcaagccctaaatatgagcagctgtccactcgattccaagggtaaag 2230
Db   2253 atattagtgtctttggggttcacgagtgtgcaatgcagtgccacggcagag 2303


RESULT  17
AAC78476
ID   AAC78476 standard; cDNA; 3305 BP.
XX
AC   AAC78476;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human PRO545 (UNQ346) nucleotide sequence SEQ ID NO:73.
XX
KW   Human; secreted protein; transmembrane protein; PRO; EST; cytostatic;
KW   expressed sequence tag; detection; cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200053756-A2.
XX
PD   14-SEP-2000.
XX
PF   18-FEB-2000; 2000WO-US04341.
XX
PR   08-MAR-1999;   99WO-US05028.
PR   12-MAR-1999;   99US-0123957.
PR   29-MAR-1999;   99US-0126773.
PR   21-APR-1999;   99US-0130232.
PR   28-APR-1999;   99US-0131445.
PR   14-MAY-1999;   99US-0134287.
PR   23-JUN-1999;   99US-0141037.
PR   26-JUL-1999;   99US-0145698.
PR   29-OCT-1999;   99US-0162506.
PR   30-NOV-1999;   99WO-US28313.
PR   02-DEC-1999;   99WO-US28551.
PR   02-DEC-1999;   99WO-US28565.
PR   16-DEC-1999;   99WO-US30095.
PR   30-DEC-1999;   99WO-US31243.
PR   30-DEC-1999;   99WO-US31274.
PR   05-JAN-2000; 2000WO-US00219.
PR   06-JAN-2000; 2000WO-US00277.
PR   06-JAN-2000; 2000WO-US00376.
XX
PA   (GETH ) GENENTECH INC.
XX
PI   Ashkenazi AJ,  Baker KP,  Botstein D,  Desnoyers L,  Eaton DL;
PI   Ferrara N,  Filvaroff E,  Fong S,  Gao W,  Gerber H,  Gerritsen ME;
PI   Goddard A,  Godowski PJ,  Grimaldi CJ,  Gurney AL,  Hillan KJ;
PI   Kljavin IJ,  Kuo SS,  Napier MA,  Pan J,  Paoni NF,  Roy MA;
PI   Shelton DL,  Stewart TA,  Tumas D,  Williams PM,  Wood WI;
XX
DR   WPI; 2000-611443/58.
DR   P-PSDB; AAB44251.
XX
```

PT Novel PRO polypeptides and polynucleotides used in detection methods,
PT to target bioactive molecules to specific cells, and to modulate
PT cellular activities -
XX
PS Claim 2; Fig 29; 636pp; English.
XX
CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed
CC sequence tag) sequences which encode secreted or transmembrane PRO
CC polypeptides. The PRO polynucleotides and polypeptides have cytostatic
CC activity. The polynucleotides and polypeptides can be used for detecting
CC the presence of PRO polypeptides in samples, for linking bioactive
CC molecules to cells and for modulating biological activities of cells,
CC using the polypeptides for specific targeting. The polypeptide targeting
CC can be used to kill the target cells, e.g. for the treatment of cancers.
CC The polypeptide pairs provide specific targeting of bioactive molecules
CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
CC the isolation of the PRO polynucleotide sequences.
XX
SQ Sequence 3305 BP; 836 A; 855 C; 945 G; 669 T; 0 other;

Query Match 4.6%; Score 113.8; DB 21; Length 3305;
Best Local Similarity 47.4%; Pred. No. 9.1e-21;
Matches 479; Conservative 0; Mismatches 517; Indels 15; Gaps 4;

Qy 1235 gaggagttggtgtgaatgagtatggtcttccaatggcagtggcacaagtattatcgcaga 1294
Db 1293 ggggaattgtcatggaccattcagacaatccccttggtgcagccgtgaccctggcacatg 1352

Qy 1295 gcctggctcaaaaccttggaatccaatgggaaccttctagcagaaagccaaaatgtgact 1354
Db 1353 agctgggccacaatttcgggatgaatcatgacacactggacaggggctgtagctgtcaaa 1412

Qy 1355 gcacagaatcctggggtggctgcatcatggaggaaacaggggtgtcccattctcgaa--- 1411
Db 1413 tggcggttgagaaaggaggctgcatcatgaacgcttccaccgggtacccatttcccatgg 1472

Qy 1412 aattttcaaagtgcagcattttggagtatagagacttttttacagagaggaggtggagcct 1471
Db 1473 tgttcagcagttgcagcaggaaggacttggagaccagcctggagaaaggaatgggggtgt 1532

Qy 1472 gccttttcaacaggccaacaa------agctatttgagcccacggaatgtggaaatggat 1525
Db 1533 gcctgtttaacctgccggaagtcagggagtctttcgggggccagaagtgtgggaacagat 1592

1526 acgtggaagctggggaggagtgtgattgtggttttcatgtggaatgctatggattatgct 1585
1593 ttgtggaagaaggagaggagtgtgactgtggggagccagaggaatgtatgaatcgctgct 1652

Qy 1586 gtaagaaatgttccctctccaacggggctcactgcagcgacgggccctgctgtaacaata 1645
Db 1653 gcaatgccaccacctgtaccctgaagccggacgctgtgtgcgcacatgggctgtgctgtg 1712

Qy 1646 cctcatgtctttttcagccacgagggtatgaatgccgggatgctgtgaacgagtgtgata 1705
Db 1713 aagactgccagctgaagcctgcaggaacagcgtgcagggactccagcaactcctgtgacc 1772

Qy 1706 ttactgaatattgtactggagactctggtcagtgcccaccaaatcttcataagcaagacg 1765
Db 1773 tcccagagttctgcacaggggccagccctcactgcccagccaatgtgtacctgcacgatg 1832

Qy 1766 gatatgcatgcaatcaaaatcagggccgctgctacaatggcgagtgcaagaccagagaca 1825
Db 1833 ggcactcatgtcaggatgtggacggctactgctacaatggcatctgccagactcacgagc 1892

Qy 1826 accagtgtcagtacatctggggaacaaaggctgcagggtctgacaagttctgctatgaaa 1885
Db 1893 agcagtgtgtcacgctctggggaccaggtgctaaacctgcccctgggatctgctttgaga 1952

Qy 1886 agctgaatacagaaggcactgagaagggaaactgcgggaaggatggaaaccggtggatt- 1944
Db 1953 gagtcaattctgcaggtgatccttatggcaactgtggcaaagtctcgaagagttcctttg 2012

Qy 1945 --cagtgcagcaaacatgatgtgttctgtggattcttactctgtaccaatcttactcgag 2002
Db 2013 ccaaatgcgagatgagagatgctaaatgtggaaaaatccagtgtcaaggaggtgccagcc 2072

Qy 2003 ctccacgtattggtcaacttcagggtgagatcattccaacttccttctaccatcaaggcc 2062
Db 2073 ggccagtcattggtaccaatgccgtttccatagaaacaaacatccctctgcagcaaggag 2132

Qy 2063 gggtgattgactgcagtggtgcccatgtagttttagatgatgata---cggatgtgggct 2119
Db 2133 gccggattctgtgccgggggacccacgtgtacttgggcgatgacatgccggacccagggc 2192

Qy 2120 atgtagaagatggaacgccatgtggcccgtctatgatgtgtttagatcggaagtgcctac 2179
Db 2193 ttgtgcttgcaggcacaaagtgtgcagatggaaaaatctgcctgaatcgtcaatgtcaaa 2252

Qy 2180 aaattcaagccctaaatatgagcagctgtccactcgattccaagggtaaag 2230
Db 2253 atattagtgtctttggggttcacgagtgtgcaatgcagtgccacggcagag 2303

RESULT 18
AAT86007
ID AAT86007 standard; cDNA; 2760 BP.
XX
AC AAT86007;
XX
DT 11-MAY-1998 (first entry)
XX
DE Mouse alpha meltrin cDNA.
XX
KW Meltrin; myotube formation; muscle tissue; therapy; membrane protein;
KW myoblast; adhesion; fusion; medicinal; murine; mouse; ss.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 41..2752
FT /*tag= a
FT /product= alpha meltrin
XX
PN WO9731109-A1.
XX
PD 28-AUG-1997.
XX
PF 17-OCT-1996; 96WO-JP03017.
XX
PR 23-FEB-1996; 96JP-0061756.
XX
PA (MOCH) MOCHIDA PHARM CO LTD.
XX
PI Fujisawa A, Mizushima C, Ogawa N, Shirakawa K, Yamakawa T;
XX
DR WPI; 1997-435161/40.
DR P-PSDB; AAW25715.
XX
PT Meltrin proteins involved with myotube formation in muscle - contain
PT metallo:proteinase and disintegrin domains
XX
PS Claim 4; Fig 2; 262pp; Japanese.
XX
CC This sequence encodes a new membrane protein, alpha meltrin, which
CC participates in the adhesion and fusion of myoblasts in the course of
CC myotube formation in muscle tissue. Meltrins and their partial peptides
CC can be used in the study of myotube formation, and in medicinal
CC compositions for muscle disorders. Meltrins contain an intracellular
CC domain, a membrane-penetrating domain, a fusion peptide domain, a
CC disintegrin domain, a metalloproteinase domain, a precursor domain and a
CC cysteine rich domain although these are not shown in the specification.
CC Meltrin antagonists can be used as bone absorption inhibitors, e.g. for
CC the treatment of osteoporosis and hypercalcaemia and for prevention of
CC cancer cell infiltration.
XX

```
SQ   Sequence 2760 BP; 693 A; 776 C; 766 G; 525 T; 0 other;

  Query Match             4.4%;  Score 110.4;  DB 18;  Length 2760;
  Best Local Similarity   47.2%;  Pred. No. 7e-20;
  Matches  586;  Conservative    0;  Mismatches  626;  Indels   30;  Gaps    7;

Qy    895 aaatatttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctcttct 954
          || ||  | || ||  | |||||    ||  ||| |  || | ||| |||  |
Db    674 aagtacgtagagctggttattgtggcagacaacagagagtttcagaggcaaggaaaagac 733

Qy    955 catgcacataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttacaag 1014
          |  |   |    ||  |      ||      | | ||| |  || ||     ||||||
Db    734 ctggagaaagttaagcagcgattaatagagatcgccaatcacgttgacaagttttaca-- 791

Qy   1015 gagcagctcaacaccagggttgtcctggtggctgtagagacctggactgagaaggatcag 1074
            |  || |||| | || | || ||||| |  || ||    |||| ||| |  ||  |
Db    792 -gaccactgaacatccggatcgtgctggtaggagtggaagtgtggaatgacatcgacaaa 850

Qy   1075 attgacatcaccaccaaccctgtgcagatgctccatgagtttctcaaaataccggcagcgc 1134
            || | |    |||| |    | ||||||||||||   | | |   |  ||
      851 tgctctataagccaggacccattcaccaggctccatgagtttctagactggagaaagata 910

     1135 attaagcagcat-------gctgatgctgtgcacctcatctcgcgggtgacatttcac 1185
          |     |  | |           ||   ||  || || ||| |||    ||||    || ||
Db    911 aagcttctacctcgaaaatcccacgacaatgctcagcttatcagtggggtttatttccaa 970

Qy   1186 tataagagaagcagtctgagttactttggaggtgtctgttct------cgcacaagagga 1239
            |  | | |  ||     | |     |  | ||  ||      |   |  ||||| 
Db    971 ggaaccaccatcggcatggcacccatcatgagcatgtgcactgcagaacagtctggagga 1030

Qy   1240 gttggtgtgaatgagtatggtcttccaatggcagtggcacaagtattatcgcagagcctg 1299
          ||||   || |  | |  |     ||  | |  |  |||      ||  | ||    |||
Db   1031 gttgtcatggaccattcagacagccccttggtgccgcagtgaccttggcacatgagctg 1090

Qy   1300 gctcaaaaccttggaatccaatgggaaccttctagcagaaagccaaaatgtgactgcaca 1359
          |  || ||| | || ||  |    ||  |       ||       |  ||       |
Db   1091 ggccacaacttcgggatgaaccatgacacactggagaggggctgcagctgcagaatggcc 1150

Qy   1360 gaatcctggggtggctgcatcatggaggaaa-caggggtgtcccattctcgaaaatttttc 1418
          | |      || ||||||||||||| |      || |||  |||||||| | |   | |||
Db   1151 gcagagaaaggaggctgcatcatgaacccgtccacgggttcccattccccatggtgttc 1210

Qy   1419 a--aagtgcagcattttggagtatagagacttttacagagaggaggtggagcctgcctt 1476
          |  |  |||||||    |||       |      |  |||  ||    ||    |||||
Db   1211 agcagctgcagcaggaaggacctggaggctagcctggagaagggcatggggatgtgcctc 1270

Qy   1477 ttcaacaggcca------acaaagctatttgagcccacggaatgtggaaatggatacgtg 1530
          ||||||   |||      |   ||   |||| |  |  | | |||||||||| || |||
Db   1271 ttcaacctaccagaggtcaagcaggcctttgggggccggaagtgtggaaatggctatgtg 1330

Qy   1531 gaagctggggaggagtgtgattgtggttttcatgtggaatgctatggattatgctgtaag 1590
          ||||  || || |||||||| || ||    |  | |||||||        ||||||||
Db   1331 gaagagggagaagagtgtgactgcggagaaccggaggaatgcacgaatcgctgctgtaac 1390

Qy   1591 aaatgttccctctccaacggggctcactgcagcgacgggccctgctgtaacaatacctca 1650
            ||    |      | |  |          || || | |      |  |
Db   1391 gctaccacctgtactctgaagccagatgctgtgtgcgcgcacgggcagtgctgtgaagac 1450

Qy   1651 tgtctttttcagccacgagggtatgaatgccgggatgctgtgaacgagtgtgatattact 1710
          ||||   |  |||| | |||   || |||| |||   |    |||   |||||  |  |
Db   1451 tgtcagctgaagcctccaggaactgcatgcaggggctccagcaactcctgtgacctccca 1510

Qy   1711 gaatattgtactggagactctggtcagtgcccaccaaatcttcataagcaagacggatat 1770
          ||||  || || ||     |  ||| || ||| | ||| |  |    || || ||  |
Db   1511 gaattctgcacagggactgcccctcactgtccagccaatgtgtacctacatgatggccac 1570

Qy   1771 gcatgcaatcaaaatcagggccgctgctacaatggcgagtgcaagaccagagacaaccag 1830
           | ||  |        | ||   ||||||||| |||   ||| |||||   ||  | |||
Db   1571 ccgtgtcagggcgtggatggttactgctacaacggcatctgccagacccatgagcagcag 1630

Qy   1831 tgtcagtacatctggggaacaaaggctgcagggtctgacaagttctgctatgaaaagctg 1890
          |||       |||||||| ||   |||  |  | ||       |||||| |||     |

Db   1631 tgtgtcacgctctggggaccaggtgctaaaccggctcctggcatctgctttgagcgagtc 1690

Qy   1891 aatacagaaggcactgagaagggaaactgcgggaaggatggagaccg---gtggattcag 1947
          ||  | | |||   |    | || ||||| || || ||     |  |    |    |
Db   1691 aactctgcaggagatccttatggtaactgtggcaaagactccaagagcgccttcgccaaa 1750

Qy   1948 tgcagcaaacatgatgtgttctgtggattcttactctgtaccaatcttactcgagctcca 2007
          ||          ||||     |||||     | |  |||     |  | | |    ||
Db   1751 tgtgagctgagagatgccaagtgtgggaaaatccagtgtcaaggtggtgcaagccgacct 1810

Qy   2008 cgtattggtcaacttcagggtgagatcattccaacttccttctaccatcaaggccgggtg 2067
            ||||||     |   | |   ||     ||| |  |    | ||  |||| |   |
Db   1811 gtcattggtaccaatgctgtttccatagaaacaaatatcccacagcaggaaggaggtcgg 1870

Qy   2068 attgactgcagtggtgcccatgtagttttagatgatgatacg 2109
          |||   ||| | ||  |||||||    || | ||||||  | |
Db   1871 attctgtgccgggggacccatgtgtacttgggtgatgacatg 1912


RESULT  19
AAT62559
ID   AAT62559 standard; DNA; 2253 BP.
XX
AC   AAT62559;
XX
DT   27-MAY-1997  (first entry)
XX
DE   Human metalloproteinase gene partial coding sequence.
XX
KW   Metalloproteinase; infection; inflammation; allergy; therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO9709420-A2.
XX
PD   13-MAR-1997.
XX
PF   05-SEP-1996;   96WO-GB02180.
XX
PR   11-JUN-1996;   96GB-0012145.
PR   05-SEP-1995;   95GB-0018026.
PR   20-OCT-1995;   95GB-0021496.
PR   20-OCT-1995;   95GB-0021497.
PR   21-DEC-1995;   95GB-0026230.
XX
PA   (CLLT ) CELLTECH THERAPEUTICS LTD.
XX
PI   Docherty AJP,  Slocombe PM;
XX
DR   WPI; 1997-192896/17.
DR   P-PSDB; AAW01825.
XX
PT   DNA encoding human metallo:proteinase - used to develop prods. for
PT   treating, e.g. infectious diseases, inflammatory disease or allergy
XX
PS   Claim 2; Page 4-6; 17pp; English.
XX
CC   A DNA fragment (AAT62559) comprises a partial coding sequence for a
CC   novel human metalloproteinase (AAW01825).  The DNA (see also AAT62558)
CC   can be obtd. by probing a human genomic or cDNA library with
CC   labelled oligonucleotide probes.  It can be used to produce the
CC   metalloproteinase in transformed prokaryotic or eukaryotic host
CC   cells; insertion downstream of the human cytomegalovirus promoter
CC   in pEE12 plasmid vectors allows expression in CHO and NSO mouse
CC   melanoma cells.  The DNA can also be used in diagnosis, e.g.
CC   to determine enzyme deficiency in a subject, or to design antisense
CC   constructs useful for the treatment of diseases or disorders
CC   related to over- or under-regulated prodn. of metalloproteinase.
XX
SQ   Sequence 2253 BP; 507 A; 656 C; 647 G; 443 T; 0 other;

  Query Match             4.3%;  Score 107.6;  DB 18;  Length 2253;
```

```
  Best Local Similarity   45.2%;  Pred. No. 3.7e-19;
  Matches  650;  Conservative    0;  Mismatches  764;  Indels   24;  Gaps    6;

Qy     892 atgaaatatttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctct 951
           ||||| ||| |||| |||    | ||   |||| |    |  || |  |||| ||||
Db     100 atgaagtatgtggagctttacctcgtggctgattatttagagtttcagaagaatcgacga 159

Qy     952 tctcatgcacataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttac 1011
             || |     |||||  ||       |       | | ||||  ||  |||    |||||
Db     160 gaccaggacgccaccaaacacaagctcatagagatcgccaactatgttgataagttttac 219

Qy    1012 aaggagcagctcaacaccagggttgtcctgg---tggctgtagagacctggactgagaag 1068
              |  || |   |  | | || |||   |||  ||   ||||       | | |
Db     220 cgatccttgaacatccggattgctctcgtgggcttggaagtgtggacccacgggaacatg 279

Qy    1069 gatcagattgacatcaccaccaaccctgtgcagatgctccatgagttctcaaaatataccgg 1128
           | |  ||      |   |  |  ||   |    |  |  |       |  |    |  |
Db     280 tgtgaagtttcagagaatccatattctaccctctggtccttttctcagttggaggcgcaag 339

Qy    1129 cagcgcattaagcagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactat 1188
           | ||       || || |   |||    |  ||  | ||| || |  || | || |||
Db     340 ctgcttgcccagaagtaccatgacaacgcccaattaatcacgggcatgtccttccacggc 399

Qy    1189 aagagaagcagtctgagttactttggaggtgtctgttct------cccacaagaggagtt 1242
           |  |  | | | |||      | |    |   | || |||       |   |  |||||||
Db     400 accaccatcggcctggccccctcatggccatgtgctctgtgtaccagtctggaggagtc 459

Qy    1243 ggtgtgaatgagtatggtcttccaatggcagtggcacaagtattatcgcagagcctggct 1302
             || |  | |   |    | | || |  |||||         |  | ||    |||
Db     460 aacatggaccactccgagaatgccattggcgtggctgccaccatggcccacgagatgggc 519

Qy    1303 caaaaccttggaatccaatgggaaccttctagcagaaagccaaaatgtgactgcacagaa 1362
           || ||| |||| ||       ||  || |      | |       |  | | |  | |
Db     520 cacaactttggcatgacccatgattctgc------agattgctgctcggccagtgcggct 573

Qy    1363 tcctggggtggctgcatcatggaggaaacaggggtgtcccattctcgaaaattttcaaag 1422
             | ||  ||  |||    |  |  || |||    ||  | |   |   ||
Db     574 gatggtgggtgcatcatggcagctgccactgggcacccctttcccaaagtgttcaatgga 633

Qy    1423 tgcagcattttggagtatagagacttttacagagaggaggtggagcctgccttttcaac 1482
           |||| ||    ||||        | | | |||  ||| ||||||   || || | ||||
Db     634 tgcaacaggagggagctggacaggtatctgcagtcaggtggtggaatgtgtctctccaac 693

Qy    1483 aggcca------acaaagctatttgagcccacggaatgtggaaatggatacgtggaagct 1536
           | ||||      |  | | | | ||    |  |   ||||| || || ||  |||||| |
Db     694 atgccagacaccaggatgttgtatggaggccggaggtgtgggaacgggtatctggaagat 753

Qy    1537 ggggaggagtgtgattgtggttttcatgtggaatgctatggattatgctgtaagaaatgt 1596
           ||||| |||||||| |||||      | | ||||||  |       ||||| ||    | |
Db     754 ggggaagagtgtgactgtggagaagaagaggaatgtaacaacccctgctgcaatgcctct 813

Qy    1597 tccctctccaacggggctcactgcagcgacgggccctgctgtaacaatacctcatgtctt 1656
                ||    || |                 |  | | |||    |     |   |||
Db     814 aattgtaccctgaggccgggggcggagtgtgctcacggctcctgctgccaccagtgtaag 873

Qy    1657 tttcagccacgagggtatgaatgccgggatgctgtgaacgagtgtgatattactgaatat 1716
            |   | | |  |||      ||||| ||    |  |   |||||||  |  | || |
Db     874 ctgttggctcctgggaccctgtgccgcgagcaggccaggcagtgtgacctcccggagttc 933

Qy    1717 tgtactggagactctggtcagtgcccaccaaatcttcataagcaagacggatatgcatgc 1776
           ||||| ||  | |||   || |||||  | ||  |  |  ||   || ||      | ||
Db     934 tgtacgggcaagtctccccactgccctaccaacttctaccagatggatggtaccccctgt 993

Qy    1777 aatcaaaatcagggccgctgctacaatggcgagtgcaagaccagagacaaccagtgtcag 1836
            |      |||| |  |||||||||| |||  ||||   |||    |  | ||||| |||
Db     994 gagggcggccaggcctactgctacaacggcatgtgcctcacctaccaggagcagtgccag 1053

Qy    1837 tacatctggggaacaaaggctgcagggtctgacaagttctgctatgaaaagctgaatgca 1896
            |  | ||||||  |    ||   |    |      |  ||||||  || ||| |||||
Db    1054 cagctgtggggacccggagcccgacctgcccctgacctctgcttcgaggaaggtgaatgtg 1113

Qy    1897 gaaggcactgagaagggaaactgcgggaaggatggagaccggtggattcagtgcagcaaa 1956
           | |||          |||||||| || |||||    ||  ||||  |  ||,       ||

Db    1114 gcaggagacacctttggaaactgtggaaagga-catgaatggtgaacacaggaagtgcaa 1172

Qy    1957 catgatgtgttctgtggattcttactctgtaccaatcttactcgagctccacgtattggt 2016
           |||||    | |    |  |    |                     |||  ||      |||
Db    1173 catgagagatgcgaagtgtgggaagatccagtgtcagagctctgaggcccggcccctgga 1232

Qy    2017 caacttcagggtgagatcattccaacttccttctaccatcaaggccgggtgattgactgc 2076
              |  | ||||    ||||   ||  |  ||  |    | ||  ||  |||  | |||
Db    1233 gtccaacgcggtgc--ccattgacaccactatcatcatgaatgggaggcagatccagtgc 1290

Qy    2077 agtggtgcccatgtagttttagatgatgatacggatgtgggctatgtagaagatggaacg 2136
             | ||  |||| ||      || |  |||   ||| |  | |    | ||   ||  |
Db    1291 cggggcacccacgtctaccgaggtcctgaggaggagggtgacatgctggacccagggctg 1350

Qy    2137 ccatgtggcccgtctatgatgtgtttagatcggaagtgcctacaaattcaagccctaaat 2196
                       | | |    | |   | |  | ||| |  |    ||   |   ||
Db    1351 gtgatgactggaaccaagtgtggctacaaccatatttgctttgaggggcagtgcaggaac 1410

Qy    2197 atgagcagctgtccactcgattccaagggtaaagtctgttcgggccatggggtgtgtagt 2256
           |    |  || |  |    ||   |   || ||    ||    ||||||||||| ||||
Db    1411 acctccttctttgaaactgaaggctgtgggaagaagtgcaatggccatggggtctgtaac 1470

Qy    2257 aatgaagccacctgcatttgtgatttcacctgggcagggacagattgcagtatccggg 2314
           ||  |    | ||||   ||         ||||||   | |    ||||  |  | ||
Db    1471 aacaaccagaactgccactgcctgccgggctgggccccgcccttctgcaacacaccgg 1528


RESULT  20
AAT62558
ID   AAT62558 standard; DNA; 2648 BP.
XX
AC   AAT62558;
XX
DT   27-MAY-1997  (first entry)
XX
DE   Human metalloproteinase DNA.
XX
KW   Metalloproteinase; infection; inflammation; allergy; therapy; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..2256
FT                   /*tag=  a
FT   3'UTR           2257..2648
FT                   /*tag=  b
XX
PN   WO9709420-A2.
XX
PD   13-MAR-1997.
XX
PF   05-SEP-1996;   96WO-GB02180.
XX
PR   11-JUN-1996;   96GB-0012145.
PR   05-SEP-1995;   95GB-0018026.
PR   20-OCT-1995;   95GB-0021496.
PR   20-OCT-1995;   95GB-0021497.
PR   21-DEC-1995;   95GB-0026230.
XX
PA   (CLLT ) CELLTECH THERAPEUTICS LTD.
XX
PI   Docherty AJP,  Slocombe PM;
XX
DR   WPI; 1997-192896/17.
DR   P-PSDB; AAW01825.
XX
PT   DNA encoding human metallo:proteinase - used to develop prods. for
PT   treating, e.g. infectious diseases, inflammatory disease or allergy
XX
PS   Claim 1; Page 2-4; 17pp; English.
XX
CC   A DNA fragment (AAT62558) includes a partial coding sequence (see also
CC   AAT62559) for a novel human metalloproteinase (AAW01825).  The DNA can
```

```
CC   be obtd. by probing a human genomic or cDNA library with labelled
CC   oligonucleotide probes.  It can be used to produce the
CC   metalloproteinase in transformed prokaryotic or eukaryotic host
CC   cells; insertion downstream of the human cytomegalovirus promoter
CC   in pEE12 plasmid vectors allows expression in CHO and NSO mouse
CC   melanoma cells.  The DNA can also be used in diagnosis, e.g.
CC   to determine enzyme deficiency in a subject, or to design antisense
CC   constructs useful for the treatment of diseases or disorders
CC   related to over- or under-regulated prodn. of metalloproteinase.
XX
SQ   Sequence 2648 BP; 620 A; 742 C; 725 G; 561 T; 0 other;


  Query Match             4.3%;  Score 107.6;  DB 18;  Length 2648;
  Best Local Similarity   45.2%;  Pred. No. 4e-19;
  Matches  650;  Conservative    0;  Mismatches  764;  Indels   24;  Gaps    6;

Qy     892 atgaaatatttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctct 951
           ||||| ||| |||| |||    | ||   |||| |   |  || | ||||| ||||
Db     102 atgaagtatgtggagctttacctcgtggctgattatttagagtttcagaagaatcgacga 161

       952 tctcatgcacataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttac 1011
             || |     |||||  ||       |       | | ||||  ||| |||    |||||
       162 gaccaggacgccaccaaacacaagctcatagagatcgccaactatgttgataagttttac 221

Qy    1012 aaggagcagctcaacaccagggttgtcctgg---tggctgtagagacctggactgagaag 1068
                 |  || |   |  | | || |||   |||  ||   ||||       | | |
Db     222 cgatccttgaacatccggattgctctcgtgggcttggaagtgtggacccacgggaacatg 281

Qy    1069 gatcagattgacatcaccaccaaccctgtgcagatgctccatgagttctcaaaataccgg 1128
           | |  ||      |   |  |  ||   |    |  | |      |  |    |  |
Db     282 tgtgaagtttcagagaatccatattctaccctctggtcctttctcagttggaggcgcaag 341

Qy    1129 cagcgcattaagcagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactat 1188
           | ||      || || |   |||    |  ||  | ||| || |  || | || |||
Db     342 ctgcttgcccagaagtaccatgacaacgcccaattaatcacgggcatgtccttccacggc 401

Qy    1189 aagagaagcagtctgagttactttggaggtgtctgttct------cgcacaagaggagtt 1242
           |  |  | | | |||     | |    |   | || |||      |   |  |||||||
Db     402 accaccatcggcctggccccctcatggccatgtgctctgtgtaccagtctggaggagtc 461

Qy    1243 ggtgtgaatgagtatggtcttccaatggcagtggcacaagtattatcgcagagcctggct 1302
             || |  | |  |    | | || |  |||||        |  | ||     |||
Db     462 aacatggaccactccgagaatgccattggcgtggctgccaccatggcccacgagatgggc 521

Qy    1303 caaaaccttggaatccaatgggaaccttctagcagaaagccaaaatgtgactgcacagaa 1362
           || ||| |||| ||       ||  || |      | |       |  | | |  | |
Db     522 cacaactttggcatgacccatgattctgc------agattgctgctcggccagtgcggct 575

Qy    1363 tcctggggtggctgcatcatggaggaaacaggggtgtcccattctcgaaaattttcaaag 1422
             | ||  ||  |||    |  |  || |||    ||  | |   |   ||
Db     576 gatggtgggtgcatcatggcagctgccactgggcacccctttcccaaagtgttcaatgga 635

       423 tgcagcattttggagtatagagacttttacagagaggaggtggagcctgccttttcaac 1482
           ||||  ||    ||||         | | | |||  ||| ||||||   || || | ||||
Db     636 tgcaacaggagggagctggacaggtatctgcagtcaggtggtggaatgtgtctctccaac 695

Qy    1483 aggcca------acaaagctatttgagcccacggaatgtggaaatggatacgtggaagct 1536
           | ||||      |  | | | | ||    |  |   ||||| || || ||  |||||| |
Db     696 atgccagacaccaggatgttgtatggaggccggaggtgtgggaacgggtatctggaagat 755

Qy    1537 ggggaggagtgtgattgtggttttcatgtggaatgctatggattatgctgtaagaaatgt 1596
           |||||  |||||||| |||||     | | ||||||  |       |||||  ||    | |
Db     756 ggggaagagtgtgactgtggagaagaagaggaatgtaacaacccctgctgcaatgcctct 815

Qy    1597 tccctctccaacggggctcactgcagcgacgggccctgctgtaacaatacctcatgtctt 1656
                ||    || |              |  | | |||    |     |   |||
Db     816 aattgtaccctgaggccgggggcggagtgtgctcacggctcctgctgccaccagtgtaag 875

Qy    1657 tttcagccacgagggtatgaatgccgggatgctgtgaacgagtgtgatattactgaatat 1716
             |   | | |  |||      |||||  ||    |  |   |||||||  |  | || |
Db     876 ctgttggctcctgggaccctgtgccgcgagcaggccaggcagtgtgacctcccggagttc 935

Qy    1717 tgtactggagactctggtcagtgcccaccaaatcttcataagcaagacggatatgcatgc 1776
           |||||  ||  | |||   || |||||  | ||  |  |  ||   || ||     | ||
Db     936 tgtacgggcaagtctccccactgccctaccaacttctaccagatggatggtacccccctgt 995

Qy    1777 aatcaaaatcagggccgctgctacaatggcgagtgcaagaccagagacaaccagtgtcag 1836
             |       |||| |  |||||||| |||  ||||   |||    |  | |||| |||
Db     996 gagggcggccaggcctactgctacaacggcatgtgcctcacctaccaggagcagtgccag 1055

Qy    1837 tacatctggggaacaaaggctgcagggtctgacaagttctgctatgaaaagctgaataca 1896
             |  | ||||||  |    ||   |    |     |  ||||||  || ||| ||||
Db     1056 cagctgtggggacccggagcccgacctgcccctgacctctgcttcgagaaggtgaatgtg 1115

Qy    1897 gaaggcactgagaagggaaactgcgggaaggatggagaccggtggattcagtgcagcaaa 1956
           | |||          |||||||| || |||||    ||  |||| |  |||       ||
Db     1116 gcaggagacacctttggaaactgtggaaagga-catgaatggtgaacacaggaagtgcaa 1174

Qy    1957 catgatgtgttctgtggattcttactctgtaccaatcttactcgagctccacgtattggt 2016
           |||||    | |   |  |   |                     |||  ||      |||
Db     1175 catgagagatgcgaagtgtgggaagatccagtgtcagagctctgaggcccggcccctgga 1234

Qy    2017 caacttcagggtgagatcattccaacttccttctaccatcaaggccgggtgattgactgc 2076
             |  | ||||   ||||   ||  |  ||  |   | ||  ||  |||  | |||
Db     1235 gtccaacgcggtgc--ccattgacaccactatcatcatgaatgggaggcagatccagtgc 1292

Qy    2077 agtggtgcccatgtagttttagatgatgatacggatgtgggctatgtagaagatggaacg 2136
           | ||  |||| ||       || |  |||   |||  |  | |    | ||    ||   |
Db     1293 cggggcacccacgtctaccgaggtcctgaggaggagggtgacatgctggacccagggctg 1352

Qy    2137 ccatgtggcccgtctatgatgtgtttagatcggaagtgcctacaaattcaagccctaaat 2196
                   | | |   | |  | | |  ||| |  |    ||  |  ||
Db     1353 gtgatgactggaaccaagtgtggctacaaccatatttgctttgaggggcagtgcaggaac 1412

Qy    2197 atgagcagctgtccactcgattccaagggtaaagtctgttcgggccatggggtgtgtagt 2256
           |   |  || |  |   ||   |  || ||   ||   ||||||||||| ||||
Db     1413 acctccttctttgaaactgaaggctgtgggaagaagtgcaatggccatggggtctgtaac 1472

Qy    2257 aatgaagccacctgcatttgtgatttcacctgggcagggacagattgcagtatccggg 2314
           ||  |    | ||||   ||         ||||||   | |    ||||  |  | ||
Db     1473 aacaaccagaactgccactgcctgccgggctgggccccgcccttctgcaacacaccgg 1530


Search completed: August 15, 2001, 21:55:57
Job time: 5720 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2001, 19:42:37 ; Search time 3312.73 Seconds
(without alignments)
11668.293 Million cell updates/sec

Title: US-09-634-252A-2
Perfect score: 2499
Sequence: 1 atgaagccgcccggcagcag.....ctcagcaaggccccatctga 2499

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1344157 seqs, 7733874588 residues

Total number of hits satisfying chosen parameters: 2688314

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 60 summaries

Database :
1: GenEmbl:*
2: gb_ba1:*
3: gb_ba2:*
4: gb_ba3:*
5: gb_in1:*
6: gb_in2:*
7: gb_in3:*
8: gb_om:*
9: gb_ov:*
10: gb_pat1:*
11: gb_pat2:*
12: gb_ph:*
13: gb_pl1:*
14: gb_pl2:*
15: gb_pl3:*
16: gb_pl4:*
17: em_ba1:*
18: em_ba2:*
19: em_fun:*
20: em_htgo_hum:*
21: em_htgo_inv:*
22: em_htgo_rod:*
23: em_htg_hum1:*
24: em_htg_hum2:*
25: em_htg_hum3:*
26: em_htg_hum4:*
27: em_htg_hum5:*
28: em_htg_hum6:*
29: em_htg_hum7:*
30: em_htg_hum8:*
31: em_htg_inv1:*
32: em_htg_inv2:*
33: em_htg_other:*
34: em_htg_rod:*
35: em_hum1:*
36: em_hum2:*
37: em_hum3:*
38: em_hum4:*
39: em_hum5:*
40: em_hum6:*
41: em_hum7:*
42: em_in:*
43: em_om:*
44: em_or:*
45: em_ov:*
46: em_pat:*
47: em_ph:*
48: em_pl:*
49: em_ro:*
50: em_sts:*
51: em_sy:*
52: em_un:*
53: em_vi:*
54: gb_sts1:*
55: gb_sts2:*
56: gb_sts3:*
57: gb_sy:*
58: gb_un:*
59: gb_vi1:*
60: gb_vi2:*
61: gb_htg1:*
62: gb_htg2:*
63: gb_htg3:*
64: gb_htg4:*
65: gb_htg5:*
66: gb_htg6:*
67: gb_htg7:*
68: gb_htg8:*
69: gb_htg9:*
70: gb_htg10:*
71: gb_htg11:*
72: gb_htg12:*
73: gb_htg13:*
74: gb_htg14:*
75: gb_htg15:*
76: gb_htg16:*
77: gb_htg17:*
78: gb_htg18:*
79: gb_htg19:*
80: gb_htg20:*
81: gb_htg21:*
82: gb_htg22:*
83: gb_htg23:*
84: gb_htg24:*
85: gb_htg25:*
86: gb_pr1:*
87: gb_pr2:*
88: gb_pr3:*
89: gb_pr4:*
90: gb_pr5:*
91: gb_pr6:*
92: gb_pr7:*
93: gb_pr8:*
94: gb_pr9:*
95: gb_ro1:*
96: gb_ro2:*
97: gb_in4:*
98: gb_pr10:*
99: em_ba3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2499	100.0	3054	10	E26532	E26532 Novel prote
2	2499	100.0	3054	85	AB009672	AB009672 Homo sapi
3	2499	100.0	3079	93	HSA5580	AJ005580 Homo sapi
4	2020	80.8	2891	94	AB009673	AB009673 Mus muscu
5	586.8	23.5	2697	10	E26531	E26531 Novel prote
6	586.8	23.5	2786	85	AB009671	AB009671 Homo sapi
7	586.8	23.5	2796	89	AF155381	AF155381 Homo sapi
8	586.8	23.5	2805	10	E26530	E26530 Novel prote

9	586.8	23.5	2858	89	AF155382	AF155382 Homo sapi
10	586.8	23.5	3259	89	AF158637	AF158637 Homo sapi
11	586.8	23.5	3295	88	AF073291	AF073291 Homo sapi
12	576.8	23.1	2773	94	AB009674	AB009674 Mus muscu
13	576.6	23.1	3168	85	AB009675	AB009675 Homo sapi
14	575	23.0	3183	10	I25848	I25848 Sequence 8
15	575	23.0	3183	10	I43485	I43485 Sequence 8
16	575	23.0	3183	45	E10493	E10493 cDNA encodi
17	562.8	22.5	3229	94	AB009676	AB009676 Mus muscu
18	540.6	21.6	3447	8	AF032383	AF032383 Xenopus l
19	530	21.2	2913	10	I25847	I25847 Sequence 7
20	530	21.2	2913	10	I43484	I43484 Sequence 7
21	530	21.2	2913	45	E10492	E10492 cDNA encodi
22	513.2	20.5	2908	97	HUMMDC	D17390 Human mRNA
23	513.2	20.5	2923	10	I25846	I25846 Sequence 6
24	513.2	20.5	2923	10	I43483	I43483 Sequence 6
25	513.2	20.5	2923	45	E10491	E10491 cDNA encodi
26	416.2	16.7	1588	8	AF032384	AF032384 Xenopus l
27	338.4	13.5	1464	10	I25845	I25845 Sequence 5
28	338.4	13.5	1464	10	I43482	I43482 Sequence 5
29	338.4	13.5	1464	45	E10490	E10490 cDNA encodi
[redacted]	220.2	8.8	163352	87	AC009225	AC009225 Homo sapi
[redacted]	220.2	8.8	194135	64	AC017081	AC017081 Homo sapi
[redacted]	178	7.1	128817	87	AC010731	AC010731 Homo sapi
c 33	178	7.1	150972	75	AC079136	AC079136 Homo sapi
c 34	173.4	6.9	144979	64	AC016280	AC016280 Homo sapi
35	154.8	6.2	2406	10	AX082155	AX082155 Sequence
36	154.8	6.2	2439	10	AX082156	AX082156 Sequence
37	143.4	5.7	3126	8	XLU66003	U66003 Xenopus lae
38	136.6	5.5	3083	8	AF032385	AF032385 Xenopus l
39	113.8	4.6	3328	88	AF023477	AF023477 Homo sapi
40	113.8	4.6	5062	88	AF023476	AF023476 Homo sapi
41	110.4	4.4	6915	94	MUSMAB	D50411 Mus musculu
c 42	110.2	4.4	142964	63	AC013380	AC013380 Homo sapi
43	110.2	4.4	244806	64	AC016822	AC016822 Homo sapi
44	107.6	4.3	2648	9	A61271	A61271 Sequence 1
45	107.6	4.3	3295	89	AF311317	AF311317 Homo sapi
46	107.6	4.3	6488	97	HSVITBIND	Y13786 Homo sapien
47	106	4.2	2971	89	AF134707	AF134707 Homo sapi
48	106	4.2	3108	89	AF326918	AF326918 Homo sapi
49	105	4.2	1777	8	AF186368	AF186368 Atractasp
50	100.8	4.0	5254	94	AF019887	AF019887 Mus muscu
51	100.8	4.0	6352	94	MUSMBA	D50410 Mus musculu
52	99.6	4.0	2251	9	A61272	A61272 Sequence 2
53	95.6	3.8	2204	5	AF252287	AF252287 Drosophil
54	95.6	3.8	2248	8	AF051787	AF051787 Gloydius
55	95.6	3.8	2290	8	CAU21003	U21003 Crotalus at
56	94.6	3.8	3412	97	HUMORF09	D14665 Homo sapien
57	94.6	3.8	3865	97	HSU41766	U41766 Human metal
58	94.4	3.8	570	8	AF055339	AF055339 Gloydius
c 59	93.6	3.7	183439	4	AC007809	AC007809 Drosophil
c 60	93.6	3.7	232744	5	AE003705	AE003705 Drosophil

[redacted]

ALIGNMENTS

RESULT 1
E26532
LOCUS E26532 3054 bp DNA PAT 07-FEB-2001
DEFINITION Novel protein belonging to MDC gene family and DNA encoding the same.
ACCESSION E26532
VERSION E26532.1 GI:13026199
KEYWORDS JP 1999155574-A/3.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3054)
AUTHORS Koji,S.
TITLE Novel protein belonging to MDC gene family and DNA encoding the same

JOURNAL Patent: JP 1999155574-A 3 15-JUN-1999;
EISAI CO LTD
COMMENT OS Homo sapiens (human)
PN JP 1999155574-A/3
PD 15-JUN-1999
PF 01-DEC-1997 JP 1997330020
PR
PI KOJI SONE
PC C12N15/09,C07K14/47,C07K16/18,C12N1/21,C12P21/02,C12Q1/68// PC
(C12N15/09,C12R1:91),(C12N1/21,C12R1:19),(C12P21/02,C12R1:19), PC
C12N15/00,
PC (C12N15/00,C12R1:91)
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT CDS 224. .2719.
FEATURES Location/Qualifiers
source 1. .3054
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 815 a 754 c 803 g 682 t
ORIGIN

Query Match 100.0%; Score 2499; DB 10; Length 3054;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2499; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 atgaagccgcccggcagcagctcgcggcagccgcccctggcgggctgcagccttgccggc 60
Db     224 ATGAAGCCGCCCGGCAGCAGCTCGCGGCAGCCGCCCCTGGCGGGCTGCAGCCTTGCCGGC 283

Qy      61 gcttcctgcggcccccaacgcggcccgccggctcggtgcctgccagcgccccggcccgc 120
Db     284 GCTTCCTGCGGCCCCCAACGCGGCCCGCCGGCTCGGTGCCTGCCAGCGCCCCGGCCCGC 343

Qy     121 acgccgccctgccgcctgcttctcgtccttctcctgctgcctccgctcgccgcctcgtcc 180
Db     344 ACGCCGCCCTGCCGCCTGCTTCTCGTCCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTCC 403

Qy     181 cggccccgcgcctgggggctgctgcgcccagcgctccgcattggaatgaaactgcagaa 240
Db     404 CGGCCCCGCGCCTGGGGGGCTGCTGCGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAA 463

Qy     241 aaaaatttgggagtcctggcagatgaagacaatacattgcaacagaatagcagcagtaat 300
Db     464 AAAAATTTGGGAGTCCTGGCAGATGAAGACAATACATTGCAACAGAATAGCAGCAGTAAT 523

Qy     301 atcagttacagcaatgcaatgcagaaagaaatcacactgccttcaagactcatatattac 360
Db     524 ATCAGTTACAGCAATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTAC 583

Qy     361 atcaaccaagactcggaaagcccttatcacgttcttgacacaaaggcaagacaccagcaa 420
Db     584 ATCAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAA 643

Qy     421 aaacataataaggctgtccatctggcccaggcaagcttccagattgaagccttcggctcc 480
Db     644 AAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCC 703

Qy     481 aaattcattcttgacctcatactgaacaatggtttgttgtcttctgattatgtggagatt 540
Db     704 AAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATT 763

Qy     541 cactacgaaaatgggaaaccacagtactctaagggtggagagcactgttactaccatgga 600
Db     764 CACTACGAAAATGGGAAACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGA 823

Qy     601 agcatcagaggcgtcaaagactccaaggtggctctgtcaacctgcaatggacttcatggc 660
Db     824 AGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGC 883

Qy     661 atgtttgaagatgataccttcgtgtatatgatagagccactagagctggttcatgatgag 720
```

away fovlik

Lolete

```
Db    884 ATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAG 943

Qy    721 aaaagcacaggtcgaccacatataatccagaaaaccttggcaggacagtattctaagcaa 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    944 AAAAGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAA 1003

Qy    781 atgaagaatctcactatggaaagaggtgaccagtggccctttctctctgaattacagtgg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1004 ATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGG 1063

Qy    841 ttgaaaagaaggaagagagcagtgaatccatcacgtggtatatttgaagaaatgaaatat 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1064 TTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATAT 1123

Qy    901 ttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgca 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1124 TTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCA 1183

Qy    961 cataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcag 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1184 CATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAG 1243

Qy   1021 ctcaacaccagggttgtcctggtggctgtagagacctggactgagaaggatcagattgac 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1244 CTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGAC 1303

Qy   1081 atcaccaccaaccctgtgcagatgctccatgagttctcaaaataccggcagcgcattaag 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1304 ATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATTAAG 1363

Qy   1141 cagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagcagt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1364 CAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGT 1423

Qy   1201 ctgagttactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtatggt 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1424 CTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGT 1483

Qy   1261 cttccaatggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatccaa 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1484 CTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAA 1543

Qy   1321 tgggaaccttctagcagaaagccaaaatgtgactgcacagaatcctggggtggctgcatc 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
     1544 TGGGAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATC 1603

     1381 atggaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttggagtat 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1604 ATGGAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTAT 1663

Qy   1441 agagacttttacagagaggaggtggagcctgccttttcaacaggccaacaaagctattt 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1664 AGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTT 1723

Qy   1501 gagcccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgtggtttt 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1724 GAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTT 1783

Qy   1561 catgtggaatgctatggattatgctgtaagaaatgttccctctccaacgggctcactgc 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1784 CATGTGGAATGCTATGGATTATGCTGTAAGAAATGTTCCCTCTCCAACGGGGCTCACTGC 1843

Qy   1621 agcgacgggccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgc 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1844 AGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGC 1903

Qy   1681 cgggatgctgtgaacgagtgtgatattactgaatattgtactggagactctggtcagtgc 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1904 CGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGC 1963

Qy   1741 ccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctac 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db   1964 CCACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTAC 2023

Qy   1801 aatggcgagtgcaagaccagagacaaccagtgtcagtacatctggggaacaaaggctgca 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2024 AATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCA 2083

Qy   1861 gggtctgacaagttctgctatgaaaagctgaatacagaaggcactgagaagggaaactgc 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2084 GGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGC 2143

Qy   1921 gggaaggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtggattctta 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2144 GGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTA 2203

Qy   1981 ctctgtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatcattcca 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2204 CTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCA 2263

Qy   2041 acttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagttttagat 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2264 ACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGAT 2323

Qy   2101 gatgatacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatgatgtgt 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2324 GATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGT 2383

Qy   2161 ttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactcgattcc 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2384 TTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCC 2443

Qy   2221 aagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatttgtgat 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2444 AAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGAT 2503

Qy   2281 ttcacctgggcagggacagattgcagtatccgggatccagttaggaaccttcacccccc 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2504 TTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCC 2563

Qy   2341 aaggatgaaggacccaagggtcctagtgccaccaatctcataataggctccatcgctggt 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2564 AAGGATGAAGGACCCAAGGGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGGT 2623

Qy   2401 gccatcctggtagcagctattgtccttgggggcacaggctggggatttaaaaatgtcaag 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2624 GCCATCCTGGTAGCAGCTATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAAATGTCAAG 2683

Qy   2461 aagagaaggttcgatcctactcagcaaggccccatctga 2499
          |||||||||||||||||||||||||||||||||||||
Db   2684 AAGAGAAGGTTCGATCCTACTCAGCAAGGCCCCATCTGA 2722
```

RESULT 2
AB009672
LOCUS AB009672 3054 bp mRNA PRI 15-AUG-1998
DEFINITION Homo sapiens mRNA for MDC3, complete cds.
ACCESSION AB009672
VERSION AB009672.1 GI:3419877
KEYWORDS MDC3.
SOURCE Homo sapiens brain cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3054)
AUTHORS Sagane,K.
TITLE Direct Submission
JOURNAL Submitted (11-DEC-1997) to the DDBJ/EMBL/GenBank databases. Koji Sagane, Eisai Co.,Ltd., Tsukuba Research Laboratories; Tokodai 5-1-3, Tsukuba, Ibaraki 300-2635, Japan
(E-mail:k1-sagane@eisai.co.jp, Tel:+81-298-47-5813, Fax:+81-298-47-5367)
REFERENCE 2 (sites)
AUTHORS Sagane,K., Ohya,Y., Hasegawa,Y. and Tanaka,I.
TITLE Metalloproteinase-like, disintegrin-like, cysteine-rich proteins

```
          MDC2 and MDC3: novel human cellular disintegrins highly expressed
          in the brain
  JOURNAL Biochem. J. 334 (Pt 1), 93-98 (1998)
  MEDLINE 98359734
FEATURES             Location/Qualifiers
     source          1. .3054
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="brain"
     CDS             224. .2722
                     /codon_start=1
                     /product="MDC3"
                     /protein_id="BAA32351.1"
                     /db_xref="GI:3419878"
                     /translation="MKPPGSSSRQPPLAGCSLAGASCGPQRGPAGSVPASAPARTPPC
                     RLLLVLLLLPPLAASSRPRAWGAAAPSAPHWNETAEKNLGVLADEDNTLQQNSSSNIS
                     YSNAMQKEITLPSRLIYYINQDSESPYHVLDTKARHQQKHNKAVHLAQASFQIEAFGS
                     KFILDLILNNGLLSSDYVEIHYENGKPQYSKGGEHCYYHGSIRGVKDSKVALSTCNGL
                     HGMFEDDTFVYMIEPLELVHDEKSTGRPHIIQKTLAGQYSKQMKNLTMERGDQWPFLS
                     ELQWLKRRKRAVNPSRGIFEEMKYLELMIVNDHKTYKKHRSSHAHTNNFAKSVVNLVD
                     SIYKEQLNTRVVLVAVETWTEKDQIDITTNPVQMLHEFSKYRQRIKQHADAVHLISRV
                     TFHYKRSSLSYFGGVCSRTRGVGVNEYGLPMAVAQVLSQSLAQNLGIQWEPSSRKPKC
                     DCTESWGGCIMEETGVSHSRKFSKCSILEYRDFLQRGGGACLFNRPTKLFEPTECGNG
                     YVEAGEECDCGFHVECYGLCCKKCSLSNGAHCSDGPCCNNTSCLFQPRGYECRDAVNE
                     CDITEYCTGDSGQCPPNLHKQDGYACNQNQGRCYNGECKTRDNQCQYIWGTKAAGSDK
                     FCYEKLNTEGTEKGNCGKDGDRWIQCSKHDVFCGFLLCTNLTRAPRIGQLQGEIIPTS
                     FYHQGRVIDCSGAHVVLDDDTDVGYVEDGTPCGPSMMCLDRKCLQIQALNMSSCPLDS
                     KGKVCSGHGVCSNEATCICDFTWAGTDCSIRDPVRNLHPPKDEGPKGPSATNLIIGSI
                     AGAILVAAIVLGGTGWGFKNVKKRRFDPTQQGPI"
BASE COUNT      815 a    754 c    803 g    682 t
ORIGIN

  Query Match           100.0%;  Score 2499;  DB 85;  Length 3054;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 2499;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atgaagccgcccggcagcagctcgcggcagccgcccctggcgggctgcagccttgccggc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     224 ATGAAGCCGCCCGGCAGCAGCTCGCGGCAGCCGCCCCTGGCGGGCTGCAGCCTTGCCGGC 283

Qy      61 gcttcctgcggcccccaacgcggccccgccggctcggtgcctgccagcgccccggcccgc 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     284 GCTTCCTGCGGCCCCCAACGCGGCCCCGCCGGCTCGGTGCCTGCCAGCGCCCCGGCCCGC 343

Qy     121 acgccgccctgccgcctgcttctcgtccttctcctgctgcctccgctcgccgcctcgtcc 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     344 ACGCCGCCCTGCCGCCTGCTTCTCGTCCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTCC 403

Qy     181 cggccccgcgcctgggggctgctgcgcccagcgctccgcattggaatgaaactgcagaa 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     404 CGGCCCCGCGCCTGGGGGGCTGCTGCGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAA 463

Qy     241 aaaaatttgggagtcctggcagatgaagacaatacattgcaacagaatagcagcagtaat 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     464 AAAAATTTGGGAGTCCTGGCAGATGAAGACAATACATTGCAACAGAATAGCAGCAGTAAT 523

Qy     301 atcagttacagcaatgcaatgcagaaagaaatcacactgccttcaagactcatatattac 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     524 ATCAGTTACAGCAATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTAC 583

Qy     361 atcaaccaagactcggaaagcccttatcacgttcttgacacaaaggcaagacaccagcaa 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     584 ATCAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAA 643

Qy     421 aaacataataaggctgtccatctggcccaggcaagcttccagattgaagccttcggctcc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     644 AAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCC 703

Qy     481 aaattcattcttgacctcatactgaacaatggtttgttgtcttctgattatgtggagatt 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     704 AAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATT 763

Qy     541 cactacgaaaatgggaaaccacagtactctaagggtggagagcactgttactaccatgga 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     764 CACTACGAAAATGGGAAACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGA 823

Qy     601 agcatcagaggcgtcaaagactccaaggtggctctgtcaacctgcaatggacttcatggc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     824 AGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGC 883

Qy     661 atgtttgaagatgataccttcgtgtatatgatagagccactagagctggttcatgatgag 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     884 ATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAG 943

Qy     721 aaaagcacaggtcgaccacatataatccagaaaaccttggcaggacagtattctaagcaa 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     944 AAAAGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAA 1003

Qy     781 atgaagaatctcactatggaaagaggtgaccagtggccctttctctctgaattacagtgg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1004 ATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGG 1063

Qy     841 ttgaaaagaaggaagagagcagtgaatccatcacgtggtatatttgaagaaatgaaatat 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1064 TTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATAT 1123

Qy     901 ttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgca 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1124 TTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCA 1183

Qy     961 cataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcag 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1184 CATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAG 1243

Qy    1021 ctcaacaccagggttgtcctggtggctgtagagacctggactgagaaggatcagattgac 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1244 CTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGAC 1303

Qy    1081 atcaccaccaaccctgtgcagatgctccatgagttctcaaaataccggcagcgcattaag 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1304 ATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATTAAG 1363

Qy    1141 cagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagcagt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1364 CAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGT 1423

Qy    1201 ctgagttactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtatggt 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1424 CTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGT 1483

Qy    1261 cttccaatggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatccaa 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1484 CTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAA 1543

Qy    1321 tgggaaccttctagcagaaagccaaaatgtgactgcacagaatcctggggtggctgcatc 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1544 TGGGAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATC 1603

Qy    1381 atggaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttggagtat 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1604 ATGGAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTAT 1663

Qy    1441 agagacttttacagagaggaggtggagcctgccttttcaacaggccaacaaagctattt 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1664 AGAGACTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTT 1723

Qy    1501 gagcccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgtggtttt 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1724 GAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTT 1783

Qy    1561 catgtggaatgctatggattatgctgtaagaaatgttccctctccaacggggctcactgc 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1784 CATGTGGAATGCTATGGATTATGCTGTAAGAAATGTTCCCTCTCCAACGGGGCTCACTGC 1843

Qy    1621 agcgacgggccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgc 1680
```

```
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1844 AGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGC 1903

Qy    1681 cgggatgctgtgaacgagtgtgatattactgaatattgtactggagactctggtcagtgc 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1904 CGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGC 1963

Qy    1741 ccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctac 1800
           |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    1964 CCACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTAC 2023

Qy    1801 aatggcgagtgcaagaccagagacaaccagtgtcagtacatctgggggaacaaaggctgca 1860
           |||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db    2024 AATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCA 2083

Qy    1861 gggtctgacaagttctgctatgaaaagctgaatacagaaggcactgagaaagggaaactgc 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2084 GGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGC 2143

Qy    1921 gggaaggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtggattctta 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2144 GGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTA 2203

Qy    1981 ctctgtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatcattcca 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2204 CTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCA 2263

Qy    2041 acttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagttttagat 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2264 ACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGAT 2323

Qy    2101 gatgatacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatgatgtgt 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2324 GATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGT 2383

Qy    2161 ttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactcgattcc 2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2384 TTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCC 2443

Qy    2221 aagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatttgtgat 2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2444 AAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGAT 2503

Qy    2281 ttcacctgggcagggacagattgcagtatccgggatccagttaggaaccttcaccccccc 2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      2504 TTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCC 2563

      2341 aaggatgaaggacccaagggtcctagtgccaccaatctcataataggctccatcgctggt 2400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2564 AAGGATGAAGGACCCAAGGGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGGT 2623

Qy    2401 gccatcctggtagcagctattgtccttgggggcacaggctggggatttaaaaatgtcaag 2460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2624 GCCATCCTGGTAGCAGCTATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAAATGTCAAG 2683

Qy    2461 aagagaaggttcgatcctactcagcaaggccccatctga 2499
           |||||||||||||||||||||||||||||||||||||||
Db    2684 AAGAGAAGGTTCGATCCTACTCAGCAAGGCCCCATCTGA 2722
```

RESULT 3
HSA5580

LOCUS	HSA5580	3079 bp	mRNA	PRI	18-JAN-2001

```
DEFINITION  Homo sapiens mRNA for adam23 protein.
ACCESSION   AJ005580
VERSION     AJ005580.1  GI:12053562
KEYWORDS    adam23 gene; Adam23 protein; disintegrin; metalloprotease.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3079)
  AUTHORS   Cal,S. and Lopez-Otin,C.
  TITLE     Identification and characterization of ADAM 23, a novel
            metalloprotease/disintegrin protein
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3078)
  AUTHORS   Lopez-Otin,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-APR-1998) Lopez-Otin C., Bioquimica y Biologia
            Molecular, Universidad de Oviedo, C/ Julian Claveria 6. Oviedo.
            Asturias., 33006, SPAIN
  REMARK    revised by author 23-APR-1998, revised by author 26-OCT-1998
FEATURES             Location/Qualifiers
     source          1. .3079
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="brain"
     5'UTR           <1. .45
                     /gene="adam23"
     gene            1. .3079
                     /gene="adam23"
     CDS             53. .2551
                     /gene="adam23"
                     /function="disintegrin"
                     /codon_start=1
                     /product="Adam23 protein"
                     /protein_id="CAC20565.1"
                     /db_xref="GI:12053563"
                     /translation="MKPPGSSSRQPPLAGCSLAGASCGPQRGPAGSVPASAPARTPPC
                     RLLLVLLLLPPLAASSRPRAWGAAAPSAPHWNETAEKNLGVLADEDNTLQQNSSSNIS
                     YSNAMQKEITLPSRLIYYINQDSESPYHVLDTKARHQQKHNKAVHLAQASFQIEAFGS
                     KFILDLILNNGLLSSDYVEIHYENGKPQYSKGGEHCYYHGSIRGVKDSKVALSTCNGL
                     HGMFEDDTFVYMIEPLELVHDEKSTGRPHIIQKTLAGQYSKQMKNLTMERGDQWPFLS
                     ELQWLKRRKRAVNPSRGIFEEMKYLELMIVNDHKTYKKHRSSHAHTNNFAKSVVNLVD
                     SIYKEQLNTRVVLVAVETWTEKDQIDITTNPVQMLHEFSKYRQRIKQHADAVHLISRV
                     TFHYKRSSLSYFGGVCSRTRGVGVNEYGLPMAVAQVLSQSLAQNLGIQWEPSSRKPKC
                     DCTESWGGCIMEETGVSHSRKFSKCSILEYRDFLQRGGGACLFNRPTKLFEPTECGNG
                     YVEAGEECDCGFHVECYGLCCKKCSLSNGAHCSDGPCCNNTSCLFQPRGYECRDAVNE
                     CDITEYCTGDSGQCPPNLHKQDGYACNQNQGRCYNGECKTRDNQCQYIWGTKAAGSDK
                     FCYEKLNTEGTEKGNCGKDGDRWIQCSKHDVFCGFLLCTNLTRAPRIGQLQGEIIPTS
                     FYHQGRVIDCSGAHVVLDDDTDVGYVEDGTPCGPSMMCLDRKCLQIQALNMSSCPLDS
                     KGKVCSGHGVCSNEATCICDFTWAGTDCSIRDPVRNLHPPKDEGPKGPSATNLIIGSI
                     AGAILVAAIVLGGTGWGFKNVKKRRFDPTQQGPI"
     3'UTR           2552. .3079
                     /gene="adam23"
BASE COUNT      865 a    714 c    769 g    731 t
ORIGIN
```

```
  Query Match             100.0%;  Score 2499;  DB 93;  Length 3079;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 2499;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 atgaagccgcccggcagcagctcgcggcagccgcccctggcgggctgcagccttgccggc 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      53 ATGAAGCCGCCCGGCAGCAGCTCGCGGCAGCCGCCCCTGGCGGGCTGCAGCCTTGCCGGC 112

Qy      61 gcttcctgcggcccccaacgcggccccgccggctcggtgcctgccagcgccccggccccgc 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     113 GCTTCCTGCGGCCCCCAACGCGGCCCCGCCGGCTCGGTGCCTGCCAGCGCCCCGGCCCGC 172

Qy     121 acgccgccctgccgcctgcttctcgtccttctcctgctgcctccgctcgccgcctcgtcc 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173 ACGCCGCCCTGCCGCCTGCTTCTCGTCCTTCTCCTGCTGCCTCCGCTCGCCGCCTCGTCC 232

Qy     181 cggccccgcgcctgggggctgctgcgcccagcgctccgcattggaatgaaactgcagaa 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     233 CGGCCCCGCGCCTGGGGGGCTGCTGCGCCCAGCGCTCCGCATTGGAATGAAACTGCAGAA 292

Qy     241 aaaaatttgggagtcctggcagatgaagacaatacattgcaacagaatagcagcagtaat 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     293 AAAAATTTGGGAGTCCTGGCAGATGAAGACAATACATTGCAACAGAATAGCAGCAGTAAT 352

Qy     301 atcagttacagcaatgcaatgcagaaagaaatcacactgccttcaagactcatatattac 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    353 ATCAGTTACAGCAATGCAATGCAGAAAGAAATCACACTGCCTTCAAGACTCATATATTAC 412

Qy    361 atcaaccaagactcggaaagcccttatcacgttcttgacacaaaggcaagacaccagcaa 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    413 ATCAACCAAGACTCGGAAAGCCCTTATCACGTTCTTGACACAAAGGCAAGACACCAGCAA 472

Qy    421 aaacataataaggctgtccatctggcccaggcaagcttccagattgaagccttcggctcc 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    473 AAACATAATAAGGCTGTCCATCTGGCCCAGGCAAGCTTCCAGATTGAAGCCTTCGGCTCC 532

Qy    481 aaattcattcttgacctcatactgaacaatggtttgttgtcttctgattatgtggagatt 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    533 AAATTCATTCTTGACCTCATACTGAACAATGGTTTGTTGTCTTCTGATTATGTGGAGATT 592

Qy    541 cactacgaaaatgggaaaccacagtactctaagggtggagagcactgttactaccatgga 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    593 CACTACGAAAATGGGAAACCACAGTACTCTAAGGGTGGAGAGCACTGTTACTACCATGGA 652

Qy    601 agcatcagaggcgtcaaagactccaaggtggctctgtcaacctgcaatggacttcatggc 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    653 AGCATCAGAGGCGTCAAAGACTCCAAGGTGGCTCTGTCAACCTGCAATGGACTTCATGGC 712

[redacted] 661 atgtttgaagatgataccttcgtgtatatgatagagccactagagctggttcatgatgag 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
[redacted] 713 ATGTTTGAAGATGATACCTTCGTGTATATGATAGAGCCACTAGAGCTGGTTCATGATGAG 772

Qy    721 aaaagcacaggtcgaccacatataatccagaaaaccttggcaggacagtattctaagcaa 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    773 AAAAGCACAGGTCGACCACATATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAGCAA 832

Qy    781 atgaagaatctcactatggaaagaggtgaccagtggccctttctctctgaattacagtgg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    833 ATGAAGAATCTCACTATGGAAAGAGGTGACCAGTGGCCCTTTCTCTCTGAATTACAGTGG 892

Qy    841 ttgaaaagaaggaagagagcagtgaatccatcacgtggtatatttgaagaaatgaaatat 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    893 TTGAAAAGAAGGAAGAGAGCAGTGAATCCATCACGTGGTATATTTGAAGAAATGAAATAT 952

Qy    901 ttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgca 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    953 TTGGAACTTATGATTGTTAATGATCACAAAACGTATAAGAAGCATCGCTCTTCTCATGCA 1012

Qy    961 cataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcag 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1013 CATACCAACAACTTTGCAAAGTCCGTGGTCAACCTTGTGGATTCTATTTACAAGGAGCAG 1072

Qy   1021 ctcaacaccagggttgtcctggtggctgtagagacctggactgagaaggatcagattgac 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1073 CTCAACACCAGGGTTGTCCTGGTGGCTGTAGAGACCTGGACTGAGAAGGATCAGATTGAC 1132

Qy   1081 atcaccaccaaccctgtgcagatgctccatgagttctcaaaataccggcagcgcattaag 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
[redacted] 1133 ATCACCACCAACCCTGTGCAGATGCTCCATGAGTTCTCAAAATACCGGCAGCGCATTAAG 1192

[redacted] 1141 cagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagcagt 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1193 CAGCATGCTGATGCTGTGCACCTCATCTCGCGGGTGACATTTCACTATAAGAGAAGCAGT 1252

Qy   1201 ctgagttactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtatggt 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1253 CTGAGTTACTTTGGAGGTGTCTGTTCTCGCACAAGAGGAGTTGGTGTGAATGAGTATGGT 1312

Qy   1261 cttccaatggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatccaa 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1313 CTTCCAATGGCAGTGGCACAAGTATTATCGCAGAGCCTGGCTCAAAACCTTGGAATCCAA 1372

Qy   1321 tgggaaccttctagcagaaagccaaaatgtgactgcacagaatcctggggtggctgcatc 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1373 TGGGAACCTTCTAGCAGAAAGCCAAAATGTGACTGCACAGAATCCTGGGGTGGCTGCATC 1432

Qy   1381 atggaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttggagtat 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1433 ATGGAGGAAACAGGGGTGTCCCATTCTCGAAAATTTTCAAAGTGCAGCATTTTGGAGTAT 1492

Qy   1441 agagacttttttacagagaggaggtggagcctgccttttcaacaggccaacaaagctattt 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1493 AGAGACTTTTTACAGAGAGGAGGTGGAGCCTGCCTTTTCAACAGGCCAACAAAGCTATTT 1552

Qy   1501 gagcccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgtggtttt 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1553 GAGCCCACGGAATGTGGAAATGGATACGTGGAAGCTGGGGAGGAGTGTGATTGTGGTTTT 1612

Qy   1561 catgtggaatgctatggattatgctgtaagaaatgttccctctccaacggggctcactgc 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1613 CATGTGGAATGCTATGGATTATGCTGTAAGAAATGTTCCCTCTCCAACGGGGCTCACTGC 1672

Qy   1621 agcgacgggccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgc 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1673 AGCGACGGGCCCTGCTGTAACAATACCTCATGTCTTTTTCAGCCACGAGGGTATGAATGC 1732

Qy   1681 cgggatgctgtgaacgagtgtgatattactgaatattgtactggagactctggtcagtgc 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1733 CGGGATGCTGTGAACGAGTGTGATATTACTGAATATTGTACTGGAGACTCTGGTCAGTGC 1792

Qy   1741 ccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctac 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1793 CCACCAAATCTTCATAAGCAAGACGGATATGCATGCAATCAAAATCAGGGCCGCTGCTAC 1852

Qy   1801 aatggcgagtgcaagaccagagacaaccagtgtcagtacatctggggaacaaaggctgca 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1853 AATGGCGAGTGCAAGACCAGAGACAACCAGTGTCAGTACATCTGGGGAACAAAGGCTGCA 1912

Qy   1861 gggtctgacaagttctgctatgaaaagctgaatacagaaggcactgagaagggaaactgc 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1913 GGGTCTGACAAGTTCTGCTATGAAAAGCTGAATACAGAAGGCACTGAGAAGGGAAACTGC 1972

Qy   1921 gggaaggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtggattctta 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1973 GGGAAGGATGGAGACCGGTGGATTCAGTGCAGCAAACATGATGTGTTCTGTGGATTCTTA 2032

Qy   1981 ctctgtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatcattcca 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2033 CTCTGTACCAATCTTACTCGAGCTCCACGTATTGGTCAACTTCAGGGTGAGATCATTCCA 2092

Qy   2041 acttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagttttagat 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2093 ACTTCCTTCTACCATCAAGGCCGGGTGATTGACTGCAGTGGTGCCCATGTAGTTTTAGAT 2152

Qy   2101 gatgatacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatgatgtgt 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2153 GATGATACGGATGTGGGCTATGTAGAAGATGGAACGCCATGTGGCCCGTCTATGATGTGT 2212

Qy   2161 ttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactcgattcc 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2213 TTAGATCGGAAGTGCCTACAAATTCAAGCCCTAAATATGAGCAGCTGTCCACTCGATTCC 2272

Qy   2221 aagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatttgtgat 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2273 AAGGGTAAAGTCTGTTCGGGCCATGGGGTGTGTAGTAATGAAGCCACCTGCATTTGTGAT 2332

Qy   2281 ttcacctgggcagggacagattgcagtatccgggatccagttaggaaccttcaccccccc 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2333 TTCACCTGGGCAGGGACAGATTGCAGTATCCGGGATCCAGTTAGGAACCTTCACCCCCCC 2392

Qy   2341 aaggatgaaggacccaagggtcctagtgccaccaatctcataataggctccatcgctggt 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2393 AAGGATGAAGGACCCAAGGGTCCTAGTGCCACCAATCTCATAATAGGCTCCATCGCTGGT 2452

Qy   2401 gccatcctggtagcagctattgtccttgggggcacaggctggggatttaaaaatgtcaag 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2453 GCCATCCTGGTAGCAGCTATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAAATGTCAAG 2512

Qy   2461 aagagaaggttcgatcctactcagcaaggccccatctga 2499
          |||||||||||||||||||||||||||||||||||||||
Db   2513 AAGAGAAGGTTCGATCCTACTCAGCAAGGCCCCATCTGA 2551
```

too late

```
RESULT   4
AB009673
LOCUS       AB009673     2891 bp    mRNA            ROD       17-AUG-1999
DEFINITION  Mus musculus mRNA for ADAM23, complete cds.
ACCESSION   AB009673
VERSION     AB009673.1  GI:5736618
KEYWORDS    ADAM23.
SOURCE      Mus musculus brain cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Sagane,K., Yamazaki,K., Mizui,Y. and Tanaka,T.
  TITLE     Cloning and chromosomal mapping of mouse ADAM11, ADAM22 and ADAM23
  JOURNAL   Gene 236 (1), 79-86 (1999)
  MEDLINE   99365303
REFERENCE   2  (bases 1 to 2891)
  AUTHORS   Sagane,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-1997) to the DDBJ/EMBL/GenBank databases. Koji
            Sagane, Eisai Co.,Ltd., Tsukuba Research Laboratories; Tokodai
            5-1-3, Tsukuba, Ibaraki 300-2635, Japan
            (E-mail:k-sagane@hhc.eisai.co.jp, Tel:+81-298-47-5813,
            Fax:+81-298-47-5367)
COMMENT     Sequence updated (07-Aug-1999).
FEATURES             Location/Qualifiers
     source          1. .2891
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /tissue_type="brain"
     CDS             80. .2569
                     /codon_start=1
                     /product="ADAM23"
                     /protein_id="BAA83381.1"
                     /db_xref="GI:5736619"
                     /translation="MKPPGSISRRPTLTGCSLPGASCGPGRCPAGPVPARAPPCRLLL
                     VLLLLPALATSSRPRARGAAAPSAPHWNETAEKTLGVLADEDNTLQQNSSSRNTSYSS
                     AVQKEITLPSRLVYYINQDSESPYHVLDTKARHQQKHNKAVHLAQASFQIEAFGSKFI
                     LDLTLNNGLLSSDYVEIHYEDGKQMYSKGGEHCYYHGSIRGVKDSRVALSTCNGLHGM
                     FEDDTFVYMIEPLELTDDEKSTGRPHIIQKTLAGQYSKQMKNLSTDGSDQWPLLPELQ
                     WLRRRKRAVNPSRGVFEEMKYLELMIVNDHKTYKKHRSSHAHTNNFAKSVVNLVDSIY
                     KEQLNTRVVLVAVETWTEKDHIDITINPVQMLHDFSKYRQRIKQHADAVHLISRVTFH
                     YKRSSLSYFGGVCSRIRGVGVNEYGLPMAVAQVLSQSLAQNLGIQWEPSSRKPKCECI
                     ESWGGCIMEETGVSHSRKFSKCSILEYRDFLQRGGGACLFNRPTKLFEPTECGNGYVE
                     AGEECDCGFHVECYGVCCKKCSLSNGAHCSDGPCCNNTSCLFQSRGYECRDAVNSCDI
                     TEYCTGDSGQCPPNLHKQDGYSCNQNQGRCYNGECKTRDNQCQYIWGTKAAGSDKFCY
                     EKLNTEGTEKGNCGKDGDRWIPCSKHDVFCGFLLCTNLTRAPRIGQLQGEIIPTSFYH
                     QGRVIDCSGAHVVLDDDTDVGYVEDGTPCGPSMMCLDRKCLQIQALNMSSCPLDSRGK
                     VCSGHGVCSNEATCICDFTWAGTDCSIRDPVRNPNPPKDEGPKGPSATNLIIGSIAGA
                     ILVAAIVLGGTGWGFKNVKKRRFDPTQQGPI"
     polyA_site      1295
                     /note="10 a nucleotides"
BASE COUNT      763 a    747 c    763 g    618 t
ORIGIN
```

Query Match 80.8%; Score 2020; DB 94; Length 2891;
Best Local Similarity 88.6%; Pred. No. 0;
Matches 2217; Conservative 0; Mismatches 270; Indels 15; Gaps 2;

```
Qy       1 atgaagccgcccggcagcagctcgcggcagccgcccctggcgggctgcagcctttgccggc 60
           |||||||||||||||||| ||| |||| |||| ||||| |||| |||||||||| |||||
Db      80 ATGAAGCCGCCCGGCAGCATCTCCCGGCGGCCGACCCTGACGGGTTGCAGCCTTCCCGGC 139

Qy      61 gcttcctgcggccccccaacgcggccccgccggctcggtgcctgccagcgccccggcccgc 120
           || |||||||||||||   ||| ||||||||||| |||||            |||||||||
Db     140 GCCTCCTGCGGCCCCGGCCGCTGCCCCGCCGGCCCGGTG-----------CCGGCCCGC 187

Qy     121 acgccgccctgccgcctgcttctcgtccttctcctgctgcctccgctcgccgcctcgtcc 180
           ||||||||||||||||||| ||||||||||||||||||| ||| |||||||| |||| |||
Db     188 GCGCCGCCCTGCCGCCTGCTCCTCGTCCTTCTCCTGCTACCTGCGCTCGCCACCTCATCC 247

Qy     181 cggccccgcgcctgggggggctgctgcgcccagcgctccgcattggaatgaaactgcagaa 240
           |||||||| ||| ||||||| |||||||||||||||||||| ||||||||||||||||||
Db     248 CGGCCCCGTGCCCGGGGGGCCGCTGCGCCCAGCGCTCCGCACTGGAATGAAACTGCAGAA 307

Qy     241 aaaaatttgggagtcctggcagatgaagacaatacattgcaacagaat---agcagcagt 297
           ||||   |||||||||||||||||||||||| |||||||||| |||   ||||||||
Db     308 AAAACCCTGGGAGTCCTGGCAGATGAAGACAACACATTGCAACAAAATAGCAGCAGCAGA 367

Qy     298 aatatcagttacagcaatgcaatgcagaaagaaatcacactgccttcaagactcatatat 357
           |||| ||| ||||||| |||| |||| |||||||||||||||||||||||||  | |||
Db     368 AATACCAGCTACAGCAGTGCAGTGCAAAAAGAAATCACACTGCCTTCAAGACTGGTGTAT 427

Qy     358 tacatcaaccaagactcggaaagcccttatcacgttcttgacacaaaggcaagacaccag 417
           |||||||||| ||||| |||||||| ||||| ||||||||||||||| ||||||||
Db     428 TACATCAACCAGGACTCAGAAAGCCCCTATCATGTTCTTGACACAAAGGCCAGACACCAA 487

Qy     418 caaaaacataataaggctgtccatctggcccaggcaagcttccagattgaagccttcggc 477
           || ||||| |||||||||| |||||||||||||||||||||||||| ||||| ||||||
Db     488 CAGAAACACAATAAGGCTGTGCATCTGGCCCAGGCAAGCTTCCAGATCGAAGCTTTCGGC 547

Qy     478 tccaaattcattcttgacctcatactgaacaatggtttgttgtcttctgattatgtggag 537
           ||||| |||||||||||||||| |||||||||||||| | |||||||| || ||||||
Db     548 TCCAAGTTCATTCTTGACCTCACACTGAACAATGGTTTGCTATCTTCTGACTACGTGGAG 607

Qy     538 attcactacgaaaatgggaaaccacagtactctaagggtggagagcactgttactaccat 597
           || ||||| ||| | ||||| |    |||||||||||||||||||||||||||||||||||
Db     608 ATCCACTATGAAGACGGGAAGCAGATGTACTCTAAGGGTGGAGAGCACTGTTACTACCAC 667

Qy     598 ggaagcatcagaggcgtcaaagactccaaggtggctctgtcaacctgcaatggacttcat 657
           |||||||||||||||||||| || |||| |||||||| || |||||||||||||| |||
Db     668 GGAAGCATCAGAGGCGTCAAGGATTCCAGGGTGGCTCTATCGACCTGCAATGGACTCCAT 727

Qy     658 ggcatgtttgaagatgataccttcgtgtatatgatagagccactagagctggttcatgat 717
           |||||||||| ||||| ||||| |||||||||||||||| || || |||  | |||||
Db     728 GGCATGTTTGAGGATGACACCTTTGTGTATATGATAGAGCCGCTGGAACTGACTGATGAT 787

Qy     718 gagaaaagcacaggtcgaccacatataatccagaaaaccttggcaggacagtattctaag 777
           |||||||||||||| |||||||| ||||||||||||||||||||||||||||||||||||
Db     788 GAGAAAAGCACAGGCCGACCACACATAATCCAGAAAACCTTGGCAGGACAGTATTCTAAG 847

Qy     778 caaatgaagaatctcactatggaaagaggtgaccagtggccctttctctctgaattacag 837
           || |||||||||||||  |  ||  |  |||||||||||||| || ||  ||||||||||
Db     848 CAGATGAAGAATCTCAGCACAGATGGCAGTGACCAGTGGCCTTTGCTACCTGAATTACAA 907

Qy     838 tggttgaaaagaaggaagagagcagtgaatccatcacgtggtatatttgaagaaatgaaa 897
           ||| ||| |||||||||| ||||| || |||||||| |||||| | ||||||||||||||
Db     908 TGGCTGAGAAGAAGGAAAAGAGCGGTCAATCCATCTCGTGGTGTGTTTGAAGAAATGAAG 967

Qy     898 tatttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctcttctcat 957
           |||||||| |||||||||||||||||||||||||||||||||||||| ||||||||||||
Db     968 TATTTGGAGCTTATGATTGTTAATGATCACAAGACGTATAAGAAGCACCGCTCTTCTCAC 1027

Qy     958 gcacataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttacaaggag 1017
           || |||||||||||||| |||||||| |||||||||||||| ||||||||||||||||||
Db    1028 GCGCATACCAACAACTTCGCAAAGTCTGTGGTCAACCTTGTAGATTCTATTTACAAGGAA 1087

Qy    1018 cagctcaacaccagggttgtcctggtggctgtagagacctggactgagaaggatcagatt 1077
           ||||||||||||||||||||||||||||||||| |||||||||| |||||||||| |||
Db    1088 CAGCTCAACACCAGGGTTGTCCTGGTGGCTGTCGAGACCTGGACCGAGAAGGATCACATT 1147

Qy    1078 gacatcaccaccaaccctgtgcagatgctccatgagttctcaaaataccggcagcgcatt 1137
           |||||||||| ||||| |||||||||| |||| |||| || ||||||||||| ||
Db    1148 GACATCACCATCAACCCCGTGCAGATGCTACATGACTTCTCCAAGTACCGGCAGCGAATC 1207

Qy    1138 aagcagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagc 1197
           || ||||| ||||| || || |||||||||||||| |||||||| || ||||||||||||
Db    1208 AAACAGCACGCTGACGCGGTCCACCTCATCTCGCGCGTGACATTCCATTATAAGAGAAGC 1267

Qy    1198 agtctgagttactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtat 1257
           |||||||||||||||||| || |||||||| | |||||| ||||||||||||||||||||
Db    1268 AGTCTGAGTTACTTTGGAGGCGTGTGTTCTCGAATAAGAGGGGTTGGTGTGAATGAGTAT 1327

Qy    1258 ggtcttccaatggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatc 1317
```

```
Db   1328 GGTCTTCCAATGGCGGTGGCACAAGTATTATCACAGAGCCTGGCTCAAAACCTTGGAATC 1387

Qy   1318 caatgggaaccttctagcagaaagccaaaatgtgactgcacagaatcctggggtggctgc 1377
Db   1388 CAGTGGGAACCTTCGAGCAGGAAGCCAAAATGTGAATGCATAGAGTCCTGGGGCGGCTGC 1447

Qy   1378 atcatggaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttggag 1437
Db   1448 ATCATGGAAGAAACAGGGGTGTCCCACTCTCGAAAGTTCTCAAAGTGCAGCATTTTGGAG 1507

Qy   1438 tatagagacttttacagagaggaggtggagcctgccttttcaacaggccaacaaagcta 1497
Db   1508 TACAGAGACTTTTTACAGAGAGGTGGCGGAGCATGTCTTTTCAATAGGCCAACTAAGCTG 1567

Qy   1498 tttgagcccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgtggt 1557
Db   1568 TTTGAGCCCACGGAATGTGGAAATGGATATGTGGAGGCCGGGGAGGAATGCGACTGTGGT 1627

Qy   1558 tttcatgtggaatgctatggattatgctgtaagaaatgttccctctccaacggggctcac 1617
     1628 TTCCATGTGGAATGCTATGGAGTTTGCTGTAAGAAGTGTTCCGCTCTCCAATGGGGCCCAC 1687

     1618 tgcagcgacgggccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaa 1677
Db   1688 TGCAGTGACGGCCCCTGCTGTAACAACACCTCATGTCTTTTTCAGTCACGAGGGTATGAA 1747

Qy   1678 tgccgggatgctgtgaacgagtgtgatattactgaatattgtactggagactctggtcag 1737
Db   1748 TGTCGGGATGCCGTAAACAGCTGTGATATCACCGAGTACTGCACTGGAGACTCTGGCCAG 1807

Qy   1738 tgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgc 1797
Db   1808 TGCCCACCGAACCTCCATAAACAAGATGGCTATAGCTGCAATCAAAATCAGGGTCGCTGC 1867

Qy   1798 tacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctggggaacaaaggct 1857
Db   1868 TACAATGGCGAGTGCAAGACAAGGGACAATCAATGCCAGTACATCTGGGGGACAAAGGCT 1927

Qy   1858 gcagggtctgacaagttctgctatgaaaagctgaatacagaaggcactgagaagggaaac 1917
Db   1928 GCGGGGTCAGACAAGTTCTGCTATGAAAAGCTGAACACGGAAGGCACCGAGAAGGGCAAT 1987

Qy   1918 tgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtggattc 1977
Db   1988 TGTGGAAAGGATGGAGACCGGTGGATCCCGTGCAGCAAGCATGATGTGTTCTGTGGATTT 2047

Qy   1978 ttactctgtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatcatt 2037
Db   2048 CTGCTTTGCACCAATCTTACCCGAGCTCCACGTATCGGTCAACTTCAAGGAGAGATCATC 2107

Qy   2038 ccaacttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagtttta 2097
     2108 CCGACTTCCTTCTATCATCAAGGCCGAGTGATTGACTGCAGTGGTGCTCATGTAGTTTTA 2167

     2098 gatgatgatacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatgatg 2157
Db   2168 GACGATGATACAGACGTGGGTTACGTTGAAGATGGGACTCCGTGTGGCCCCTCCATGATG 2227

Qy   2158 tgtttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactcgat 2217
Db   2228 TGCTTAGATCGGAAGTGCCTACAGATTCAAGCCCTGAATATGAGCAGCTGCCCACTTGAC 2287

Qy   2218 tccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatttgt 2277
Db   2288 TCAAGGGGTAAAGTCTGCTCCGGCCACGGGGTGTGTAGCAACGAAGCCACCTGCATCTGT 2347

Qy   2278 gatttcacctgggcagggacagattgcagtatccgggatccagttaggaaccttcacccc 2337
Db   2348 GATTTCACTTGGGCAGGCACAGACTGCAGCATCCGGGATCCAGTTCGGAACCCCAACCCC 2407

Qy   2338 cccaaggatgaaggacccaagggtcctagtgccaccaatctcataataggctccatcgct 2397
```

```
Db   2408 CCTAAGGATGAAGGCCCTAAGGGTCCTAGCGCCACCAATCTCATAATAGGCTCCATCGCT 2467

Qy   2398 ggtgccatcctggtagcagctattgtccttgggggcacaggctggggatttaaaaatgtc 2457
Db   2468 GGTGCCATCCTGGTAGCAGCTATTGTCCTTGGGGGCACAGGCTGGGGATTTAAAAACGTC 2527

Qy   2458 aagaagagaaggttcgatcctactcagcaaggccccatctga 2499
Db   2528 AAGAAGAGGAGATTCGATCCCACTCAGCAAGGCCCCATCTGA 2569
```

RESULT 5
E26531
LOCUS E26531 2697 bp DNA PAT 07-FEB-2001
DEFINITION Novel protein belonging to MDC gene family and DNA encoding the same.
ACCESSION E26531
VERSION E26531.1 GI:13026198
KEYWORDS JP 1999155574-A/2.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2697)
AUTHORS Koji,S.
TITLE Novel protein belonging to MDC gene family and DNA encoding the same
JOURNAL Patent: JP 1999155574-A 2 15-JUN-1999;
EISAI CO LTD
COMMENT OS Homo sapiens (human)
PN JP 1999155574-A/2
PD 15-JUN-1999
PF 01-DEC-1997 JP 1997330020
PR
PI KOJI SONE
PC C12N15/09,C07K14/47,C07K16/18,C12N1/21,C12P21/02,C12Q1/68// PC (C12N15/09,C12R1:91),(C12N1/21,C12R1:19),(C12P21/02,C12R1:19), PC C12N15/00,
PC (C12N15/00,C12R1:91)
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT CDS 122. .2590.
FEATURES Location/Qualifiers
source 1. .2697
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 795 a 529 c 699 g 674 t
ORIGIN

Query Match 23.5%; Score 586.8; DB 10; Length 2697;
Best Local Similarity 57.2%; Pred. No. 3.4e-130;
Matches 1237; Conservative 0; Mismatches 887; Indels 38; Gaps 8;

```
Qy    320 tgcagaaagaaatcacactgccttcaagactcatatattacatcaaccaagactcggaaa 379
Db    255 TGGAGCGCCAGAGCATCGTGCCACTGCGCCTCATCTACCGCTCGGGCGGCGAAGACGAAA 314

Qy    380 gcccttatcacgttcttgacacaaaggcaagacaccagcaaaaacataataaggctgtcc 439
Db    315 GTCGGCACGACGCGCTCGACACGCGGGTGCGGGGCGACCTCGGTGGCCCGCAGTTGACTC 374

Qy    440 atctggcccaggcaagcttccagattgaagccttcggctccaaattcattcttgacctca 499
Db    375 ATGTTGACCAAGCAAGCTTCCAGGTTGATGCCTTTGGAACGTCATTCATTCTCGATGTCG 434

Qy    500 tactgaacaatggtttgttgtcttctgattatgtggagattcactacgaaaat---ggga 556
Db    435 TGCTAAATCATGATTTGCTGTCCTCTGAATACATAGAGAGACACATTGAACATGGAGGCA 494

Qy    557 aaccacagtactctaagggtggagagcactgttactaccatggaagcatcagaggcgtca 616
```

```
Db    495 AGACTGTGGAAGTTAAAGGAGGAGAGCACTGTTACTACCAGGGCCATATCCGAGGAAACC 554

Qy    617 aagactccaaggtggctctgtcaacctgcaatggacttcatggcatgtttgaagatgata 676

Db    555 CTGACTCATTTGTTGCATTGTCAACATGCCACGGACTTCATGGGATGTTCTATGACGGGA 614

Qy    677 ccttcgtgtatatgatagagccactagagctggttcatgatgagaaaagcacaggtcgac 736

Db    615 ACCACACATATCTCATTGAGCCAGAAGAAAATG----ACACTACTCAAGAGGATTTCCAT 670

Qy    737 cacatataatccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcacta 796

Db    671 TTTCATTCAGTTTACAAATCCAGACTGTTTGAATTTTCCTTGGATGATCTTCCATCTGAA 730

      797 tggaaagaggtgaccagtggccctttctctctgaattacagtggttgaaaagaaggaaga 856

      731 TTTCAGCAAGTAAACATTACTCCATCAAAATTTATTTTGAAGCCAAGACCAAAAAGGAGT 790

Qy    857 gagcagtgaatccat-------cacgtggtatatttgaagaaatgaaatatttggaactt 909

Db    791 AAACGGCAGCTTCGTCGATATCCTCGTAATGTAGAAGAAGAAACCAAATACATTGAACTG 850

Qy    910 atgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgcacataccaac 969

Db    851 ATGATTGTGAATGATCACCTTATGTTTAAAAAACATCGGCTTTCCGTTGTACATACCAAT 910

Qy    970 aactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcagctcaacacc 1029

Db    911 ACCTATGCGAAATCTGTGGTGAACATGGCAGATTTAATATATAAAGACCAACTTAAGACC 970

Qy   1030 agggttgtcctggtggctgtagagacctggactgagaaggatcagattgacatcaccacc 1089

Db    971 AGGATAGTATTGGTTGCTATGGAAACCTGGGCGACTGACAACAAGTTTGCCATATCTGAA 1030

Qy   1090 aaccctgtgcagatgctccatgagttctcaaaataccg---gcagcgcattaagcagcat 1146

Db   1031 AATCCATTGATCACCCTACGTGAGTTTATGAAATACAGGAGGGATTTTATCAAAGAGAAA 1090

Qy   1147 gctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagcagtctgagt 1206

Db   1091 AGTGATGCAGTTCACCTTTTTTCGGGAAGTCAATTTGAGAGTAGCCGGAGCGGGGCAGCT 1150

Qy   1207 tactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtatggtcttcca 1266

Db   1151 TATATTGGTGGGATTTGCTCGTTGCTGAAAGGAGGAGGCGTGAATGAATTTGGGAAAACT 1210

Qy   1267 atggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatccaatgggaa 1326

Db   1211 GATTTAATGGCTGTTACACTTGCCCAGTCATTAGCCCATAATATTGGTATTATCTCAGAC 1270

Qy   1327 ccttcta---gcagaaagccaaaatgtgactgcacagaatcctgggggtggctgcatcatg 1383

Db   1271 AAAAGAAAGTTAGCAAGTGGTGAATGTAAATGCGAGGACACGTGGTCCGGGTGCATAATG 1330

Qy   1384 gaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcatcttggagtataga 1443

Db   1331 GGAGACACTGGCTATTATCTTCCTAAAAAGTTCACCCAGTGTAATATTGAAGAGTATCAT 1390

Qy   1444 gacttttacagagaggaggtggagcctgccttttcaacaggccaacaaagctatttgag 1503

Db   1391 GACTTCCTGAATAGTGGAGGTGGTGCCTGCCTTTTCAACAAACCTTCCAAGCTTCTTGAT 1450

Qy   1504 cccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgtggttttcat 1563

Db   1451 CCTCCTGAGTGTGGCAATGGCTTCATTGAAACTGGAGAGGAGTGTGATTGTGGAACCCCG 1510

Qy   1564 gtggaatg---------ctatggattatgctgtaagaaatgttccctctccaacggggct 1614

Db   1511 GCCGAATGTGTCCTTGAAGGAGCAGAGTGTTGTAAGAAATGCACCTTGACTCAAGACTCT 1570

Qy   1615 cactgcagcgacgggccctgctgtaacaatacctcatgtctttttcagccacgagggtat 1674

Db   1571 CAATGCAGTGACGGTCTTTGCTGTAAAAA------GTGCAAGTTTCAGCCTATGGGCACT 1624
```

```
Qy   1675 gaatgccgggatgctgtgaacgagtgtgatattactgaatattgtactggagactctggt 1734

Db   1625 GTGTGCCGAGAAGCAGTAAATGATTGTGATATTCGTGAAACGTGCTCAGGAAATTCAAGC 1684

Qy   1735 cagtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatcagggccgc 1794

Db   1685 CAGTGTGCCCCTAATATTCATAAAATGGATGGATATTCATGTGATGGTGTTCAGGGAATT 1744

Qy   1795 tgctacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctggggaacaaag 1854

Db   1745 TGCTTTGGAGGAAGATGCAAAACCAGAGATAGACAATGCAAATACATTTGGGGGCAAAAG 1804

Qy   1855 gctgcagggtctgacaagttctgctatgaaaagctgaatacagaaggcactgagaaggga 1914

Db   1805 GTGACAGCATCAGACAAATATTGCTATGAGAAACTGAATATTGAAGGGACGGAGAAGGGT 1864

Qy   1915 aactgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtgga 1974

Db   1865 AACTGTGGGAAAGACAAAGACACATGGATACAGTGCAACAAACGGGATGTGCTTTGTGGT 1924

Qy   1975 ttcttactctgtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatc 2034

Db   1925 TACCTTTTGTGTACCAATATTGGCAATATCCCAAGGCTTGGAGAACTCGATGGTGAAATC 1984

Qy   2035 attccaacttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagtt 2094

Db   1985 ACATCTACTTTAGTTGTGCAGCAAGGAAGAACATTAAACTGCAGTGGTGGGCATGTTAAG 2044

Qy   2095 ttagatgatgatacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatg 2154

Db   2045 CTTGAAGAAGATGTAGATCTTGGCTATGTGGAAGATGGGACACCTTGTGGTCCCCAAATG 2104

Qy   2155 atgtgtttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactc 2214

Db   2105 ATGTGCTTAGAACACAGGTGTCTTCCTGTGGCTTCTTTCAACTTTAGTACTTGCTTGAGC 2164

Qy   2215 gattccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatt 2274

Db   2165 AGTAAAGAAGGCACTATTTGCTCAGGAAATGGAGTTTGCAGTAATGAGCTGAAGTGTGTG 2224

Qy   2275 tgtgatttcacctgggcagggacagattg---cagtatccgggatccagttaggaacctt 2331

Db   2225 TGTAACAGACACTGGATAGGTTCTGATTGCAACACTTACTTCCCTCACAATGATGATGCA 2284

Qy   2332 caccccccaaggatgaaggacccaagggtcctagtgccaccaatctcataataggctcc 2391

Db   2285 AAGACTGGTATCACTCTGTCTGGCAATGGTGTTGCTGGCACCAATATCATAATAGGCATA 2344

Qy   2392 atcgctggtgccatcctggtagcagctattgtccttgggggcacaggctggggatttaaa 2451

Db   2345 ATTGCTGGCACCATTTTAGTGCTGGCCCTCATATTAGGAATAACTGCGTGGGGTTATAAA 2404

Qy   2452 aa 2453
          ||
Db   2405 AA 2406
```

```
RESULT   6
AB009671
LOCUS       AB009671     2786 bp    mRNA            PRI       15-AUG-1998
DEFINITION  Homo sapiens mRNA for MDC2 alpha, MDC2 beta, complete cds.
ACCESSION   AB009671
VERSION     AB009671.1  GI:3419874
KEYWORDS    MDC2 beta; MDC2 alpha.
SOURCE      Homo sapiens brain cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2786)
  AUTHORS   Sagane,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-1997) to the DDBJ/EMBL/GenBank databases. Koji
```

```
            Sagane, Eisai Co.,Ltd., Tsukuba Research Laboratories; Tokodai
            5-1-3, Tsukuba, Ibaraki 300-2635, Japan
            (E-mail:kl-sagane@eisai.co.jp, Tel:+81-298-47-5813,
            Fax:+81-298-47-5367)
REFERENCE   2  (sites)
  AUTHORS   Sagane,K., Ohya,Y., Hasegawa,Y. and Tanaka,I.
  TITLE     Metalloproteinase-like, disintegrin-like, cysteine-rich proteins
            MDC2 and MDC3: novel human cellular disintegrins highly expressed
            in the brain
  JOURNAL   Biochem. J. 334 (Pt 1), 93-98 (1998)
  MEDLINE   98359734
FEATURES              Location/Qualifiers
     source           1. .2786
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /tissue_type="brain"
     CDS              join(122. .2421,2530. .2701)
                      /codon_start=1
                      /product="MDC2 beta"
                      /protein_id="BAA32350.1"
                      /db_xref="GI:3419876"
                      /translation="MQAAVAVSVPFLLLCVLGTCPPARCGQAGDASLMELEKRKENRF
                      VERQSIVPLRLIYRSGGEDESRHDALDTRVRGDLGGPQLTHVDQASFQVDAFGTSFIL
                      DVVLNHDLLSSEYIERHIEHGGKTVEVKGGEHCYYQGHIRGNPDSFVALSTCHGLHGM
                      FYDGNHTYLIEPEENDTTQEDFHFHSVYKSRLFEFSLDDLPSEFQQVNITPSKFILKP
                      RPKRSKRQLRRYPRNVEEETKYIELMIVNDHLMFKKHRLSVVHTNTYAKSVVNMADLI
                      YKDQLKTRIVLVAMETWATDNKFAISENPLITLREFMKYRRDFIKEKSDAVHLFSGSQ
                      FESSRSGAAYIGGICSLLKGGGVNEFGKTDLMAVTLAQSLAHNIGIISDKRKLASGEC
                      KCEDTWSGCIMGDTGYYLPKKFTQCNIEEYHDFLNSGGGACLFNKPSKLLDPPECGNG
                      FIETGEECDCGTPAECVLEGAECCKKCTLTQDSQCSDGLCCKKCKFQPMGTVCREAVN
                      DCDIRETCSGNSSQCAPNIHKMDGYSCDGVQGICFGGRCKTRDRQCKYIWGQKVTASD
                      KYCYEKLNIEGTEKGNCGKDKDTWIQCNKRDVLCGYLLCTNIGNIPRLGELDGEITST
                      LVVQQGRTLNCSGGHVKLEEDVDLGYVEDGTPCGPQMMCLEHRCLPVASFNFSTCLSS
                      KEGTICSGNGVCSNELKCVCNRHWIGSDCNTYFPHNDDAKTGITLSGNGVAGTNIIIG
                      IIAGTILVLALILGITAWGYKNYREQRSNGLSHSWSERIPDTKHISDICENGRPRSNS
                      WQGNLGGNKKKIRGKRFRPRSNSTE"
     CDS              122. .2701
                      /codon_start=1
                      /product="MDC2 alpha"
                      /protein_id="BAA32349.1"
                      /db_xref="GI:3419875"
                      /translation="MQAAVAVSVPFLLLCVLGTCPPARCGQAGDASLMELEKRKENRF
                      VERQSIVPLRLIYRSGGEDESRHDALDTRVRGDLGGPQLTHVDQASFQVDAFGTSFIL
                      DVVLNHDLLSSEYIERHIEHGGKTVEVKGGEHCYYQGHIRGNPDSFVALSTCHGLHGM
                      FYDGNHTYLIEPEENDTTQEDFHFHSVYKSRLFEFSLDDLPSEFQQVNITPSKFILKP
                      RPKRSKRQLRRYPRNVEEETKYIELMIVNDHLMFKKHRLSVVHTNTYAKSVVNMADLI
                      YKDQLKTRIVLVAMETWATDNKFAISENPLITLREFMKYRRDFIKEKSDAVHLFSGSQ
                      FESSRSGAAYIGGICSLLKGGGVNEFGKTDLMAVTLAQSLAHNIGIISDKRKLASGEC
                      KCEDTWSGCIMGDTGYYLPKKFTQCNIEEYHDFLNSGGGACLFNKPSKLLDPPECGNG
                      FIETGEECDCGTPAECVLEGAECCKKCTLTQDSQCSDGLCCKKCKFQPMGTVCREAVN
                      DCDIRETCSGNSSQCAPNIHKMDGYSCDGVQGICFGGRCKTRDRQCKYIWGQKVTASD
                      KYCYEKLNIEGTEKGNCGKDKDTWIQCNKRDVLCGYLLCTNIGNIPRLGELDGEITST
                      LVVQQGRTLNCSGGHVKLEEDVDLGYVEDGTPCGPQMMCLEHRCLPVASFNFSTCLSS
                      KEGTICSGNGVCSNELKCVCNRHWIGSDCNTYFPHNDDAKTGITLSGNGVAGTNIIIG
                      IIAGTILVLALILGITAWGYKNYREQRQLPQGDYVKKPGDGDSFYSDIPPGVSTNSAS
                      SSKKRSNGLSHSWSERIPDTKHISDICENGRPRSNSWQGNLGGNKKKIRGKRFRPRSN
                      STE"
     polyA_site       2786
                      /note="19 a nucleotides"
BASE COUNT      810 a    553 c    723 g    700 t
ORIGIN


  Query Match           23.5%;  Score 586.8;  DB 85;  Length 2786;
  Best Local Similarity   57.2%;  Pred. No. 3.4e-130;
  Matches 1237;  Conservative    0;  Mismatches  887;  Indels   38;  Gaps    8;

Qy    320 tgcagaaagaaatcacactgccttcaagactcatatattacatcaaccaagactcggaaa 379
          || ||    | | ||   ||||     | ||||| ||   |     |   ||    ||||
Db    255 TGGAGCGCCAGAGCATCGTGCCACTGCGCCTCATCTACCGCTCGGGCGGCGAAGACGAAA 314

Qy    380 gcccttatcacgttcttgacacaaaggcaagacaccagcaaaaacataataaggctgtcc 439
          | |   |  |||  || |||||   ||   |   | | |            ||      |

Db    315 GTCGGCACGACGCGCTCGACACGCGGGTGCGGGGCGACCTCGGTGGCCCGCAGTTGACTC 374

Qy    440 atctggcccaggcaagcttccagattgaagccttcggctccaaattcattcttgacctca 499
          || | | ||| ||||||||||||| |||| ||||| ||  |   |||||||| ||  ||
Db    375 ATGTTGACCAAGCAAGCTTCCAGGTTGATGCCTTTGGAACGTCATTCATTCTCGATGTCG 434

Qy    500 tactgaacaatggtttgttgtcttctgattatgtggagattcactacgaaaat---ggga 556
          | || ||  ||| |||| |||| ||||| ||  | |||| |||  ||| || |  || |
Db    435 TGCTAAATCATGATTTGCTGTCCTCTGAATACATAGAGAGACACATTGAACATGGAGGCA 494

Qy    557 aaccacagtactctaagggtggagagcactgttactaccatggaagcatcagaggcgtca 616
          |  |   | |   ||| || ||||||||||||||||||||| ||    ||| ||||   |
Db    495 AGACTGTGGAAGTTAAAGGAGGAGAGCACTGTTACTACCAGGGCCATATCCGAGGAAACC 554

Qy    617 aagactccaaggtggctctgtcaacctgcaatggacttcatggcatgtttgaagatgata 676
            |||||    || ||  ||||||| ||| | |||||||||||| |||||  | || |  |
Db    555 CTGACTCATTTGTTGCATTGTCAACATGCCACGGACTTCATGGGATGTTCTATGACGGGA 614

Qy    677 ccttcgtgtatatgatagagccactagagctggttcatgatgagaaaagcacaggtcgac 736
          |  |   ||| | || ||||||  |||     |       |  |   |||   |  || |
Db    615 ACCACACATATCTCATTGAGCCAGAAGAAAATG----ACACTACTCAAGAGGATTTCCAT 670

Qy    737 cacatataatccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcacta 796
              | |      ||| |  | | |       ||       |          ||   |
Db    671 TTTCATTCAGTTTACAAATCCAGACTGTTTGAATTTTCCTTGGATGATCTTCCATCTGAA 730

Qy    797 tggaaagaggtgaccagtggccctttctctctgaattacagtggttgaaaagaaggaaga 856
          |   |  | || | || |   || |      | | ||  |      ||  | || | ||
Db    731 TTTCAGCAAGTAAACATTACTCCATCAAAATTTATTTTGAAGCCAAGACCAAAAAGGAGT 790

Qy    857 gagcagtgaatccat-------cacgtggtatatttgaagaaatgaaatatttggaactt 909
            | |  |    | | |       | |||  | ||   |||||||  |||||  | |||||
Db    791 AAACGGCAGCTTCGTCGATATCCTCGTAATGTAGAAGAAGAAACCAAATACATTGAACTG 850

Qy    910 atgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgcacataccaac 969
          |||||||| |||||||||   | || ||| || |||||   |||   || ||||||||||
Db    851 ATGATTGTGAATGATCACCTTATGTTTAAAAAACATCGGCTTTCCGTTGTACATACCAAT 910

Qy    970 aactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcagctcaacacc 1029
          | || ||| || || ||||| ||| | |  ||||  || || || || || || || |||
Db    911 ACCTATGCGAAATCTGTGGTGAACATGGCAGATTTAATATATAAAGACCAACTTAAGACC 970

Qy   1030 agggttgtcctggtggctgtagagacctggactgagaaggatcagattgacatcaccacc 1089
          ||| | ||  |||| ||| | || |||||  |      |  |  || ||| |||  |
Db    971 AGGATAGTATTGGTTGCTATGGAAACCTGGGCGACTGACAACAAGTTTGCCATATCTGAA 1030

Qy   1090 aaccctgtgcagatgctccatgagttctcaaaataccg---gcagcgcattaagcagcat 1146
          || ||  ||   |  || | ||||||    |||||| |   | |    || ||  || |
Db   1031 AATCCATTGATCACCCTACGTGAGTTTATGAAATACAGGAGGGATTTTATCAAAGAGAAA 1090

Qy   1147 gctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagcagtctgagt 1206
            ||||||  || |||||  | |||  |      ||||  |   ||   | |||  |     |
Db   1091 AGTGATGCAGTTCACCTTTTTTCGGGAAGTCAATTTGAGAGTAGCCGGAGCGGGGCAGCT 1150

Qy   1207 tactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtatggtcttcca 1266
          ||  |||| ||  |  || ||       | |||||  || |||||||| | |||     |
Db   1151 TATATTGGTGGGATTTGCTCGTTGCTGAAAGGAGGAGGCGTGAATGAATTTGGGAAAACT 1210

Qy   1267 atggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatccaatgggaa 1326
               | ||||      | |  | |||    | || || ||  |||| ||    | ||
Db   1211 GATTTAATGGCTGTTACACTTGCCCAGTCATTAGCCCATAATATTGGTATTATCTCAGAC 1270

Qy   1327 ccttcta---gcagaaagccaaaatgtgactgcacagaatcctggggtggctgcatcatg 1383
                |       ||      |||| | |||  ||  | |||   || ||||| |||
Db   1271 AAAAGAAAGTTAGCAAGTGGTGAATGTAAATGCGAGGACACGTGGTCCGGGTGCATAATG 1330

Qy   1384 gaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttggagtataga 1443
          |  || || ||    |  | | ||  ||| || |  ||||  |  |||   ||||||
Db   1331 GGAGACACTGGCTATTATCTTCCTAAAAAGTTCACCCAGTGTAATATTGAAGAGTATCAT 1390

Qy   1444 gacttttacagagaggaggtggagcctgccttttcaacaggccaacaaagctatttgag 1503
          |||||  |  | || |||||||| |||||||||||||||||  ||  | |||||  ||||
Db   1391 GACTTCCTGAATAGTGGAGGTGGTGCCTGCCTTTTCAACAAACCTTCTAAGCTTCTTGAT 1450
```

```
Qy   1504 cccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgtggttttcat 1563
          ||  | || ||||| ||||| | | | ||| |||| ||||||||||||||||||    |
Db   1451 CCTCCTGAGTGTGGCAATGGCTTCATTGAAACTGGAGAGGAGTGTGATTGTGGAACCCCG 1510

Qy   1564 gtggaatg---------ctatggattatgctgtaagaaatgttccctctccaacggggct 1614
          |  |||||                | |   || ||||||||||  || |  |  | |   ||
Db   1511 GCCGAATGTGTCCTTGAAGGAGCAGAGTGTTGTAAGAAATGCACCTTGACTCAAGACTCT 1570

Qy   1615 cactgcagcgacgggccctgctgtaacaatacctcatgtctttttcagccacgagggtat 1674
          || ||||| ||||| |  |||||||| ||        ||    ||||||||    ||   |
Db   1571 CAATGCAGTGACGGTCTTTGCTGTAAAAA------GTGCAAGTTTCAGCCTATGGGCACT 1624

Qy   1675 gaatgccgggatgctgtgaacgagtgtgatattactgaatattgtactggagactctggt 1734
          |  ||||| || || || || || |||||||||  ||||   ||  | ||| | ||  |
Db   1625 GTGTGCCGAGAAGCAGTAAATGATTGTGATATTCGTGAAACGTGCTCAGGAAATTCAAGC 1684

Qy   1735 cagtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatcagggccgc 1794
          |||||  | || ||| |||||||    || |||||| |||| ||     ||||||
Db   1685 CAGTGTGCCCCTAATATTCATAAAATGGATGGATATTCATGTGATGGTGTTCAGGGAATT 1744

Qy   1795 tgctacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctgggaacaaag 1854
          ||||     ||    ||||| |||||||| |  || ||  | ||||| |||||   ||||
Db   1745 TGCTTTGGAGGAAGATGCAAAACCAGAGATAGACAATGCAAATACATTTGGGGGCAAAAG 1804

Qy   1855 gctgcagggtctgacaagttctgctatgaaaagctgaatacagaaggcactgagaaggga 1914
          |  |||  || ||||| |  |||||||| || |||||||  ||||| || ||||||||
Db   1805 GTGACAGCATCAGACAAATATTGCTATGAGAAACTGAATATTGAAGGGACGGAGAAGGGT 1864

Qy   1915 aactgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtgga 1974
          ||||| ||||| ||   ||||   ||||| ||||||| |||||  |||||| | |||||
Db   1865 AACTGTGGGAAAGACAAAGACACATGGATACAGTGCAACAAACGGGAATGTGCTTTGTGGT 1924

Qy   1975 ttcttactctgtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatc 2034
          | | |  | |||||||||| ||          ||| |  ||||  ||||  | ||||| |||
Db   1925 TACCTTTTGTGTACCAATATTGGCAATATCCCAAGGCTTGGAGAACTCGATGGTGAAATC 1984

Qy   2035 attccaacttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagtt 2094
          |   | ||||   |    || |||||  |      |  ||||||||||||  |||||
Db   1985 ACATCTACTTTAGTTGTGCAGCAAGGAAGAACATTAAACTGCAGTGGTGGGCATGTTAAG 2044

Qy   2095 ttagatgatgatacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatg 2154
          | || || |||   ||| | |||||||| |||||||| || || ||||| ||    |||
Db   2045 CTTGAAGAAGATGTAGATCTTGGCTATGTGGAAGATGGGACACCTTGTGGTCCCCAAATG 2104

     2155 atgtgtttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactc 2214
          ||||| ||||| |  | ||| || |   |      |  | ||  | || |  ||      |
     2105 ATGTGCTTAGAACACAGGTGTCTTCCTGTGGCTTCTTTCAACTTTAGTACTTGCTTGAGC 2164

Qy   2215 gattccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatt 2274
          |    | || |   | || || ||  |||| || || ||||||||    |  ||  |
Db   2165 AGTAAAGAAGGCACTATTTGCTCAGGAAATGGAGTTTGCAGTAATGAGCTGAAGTGTGTG 2224

Qy   2275 tgtgatttcacctgggcagggacagattg---cagtatccgggatccagttaggaacctt 2331
          ||| |      ||||  |||  | |||||   || |  |     ||    |   |
Db   2225 TGTAACAGACACTGGATAGGTTCTGATTGCAACACTTACTTCCCTCACAATGATGATGCA 2284

Qy   2332 caccccccaaggatgaaggacccaagggtcctagtgccaccaatctcataataggctcc 2391
          |  |    |    |         ||| |||  |  || |||||| ||||||||||
Db   2285 AAGACTGGTATCACTCTGTCTGGCAATGGTGTTGCTGGCACCAATATCATAATAGGCATA 2344

Qy   2392 atcgctggtgccatcctggtagcagctattgtccttgggggcacaggctgggggatttaaa 2451
          || |||||  ||||  | ||    ||  |  |  | ||    || |  ||||| | ||||
Db   2345 ATTGCTGGCACCATTTTAGTGCTGGCCCTCATATTAGGAATAACTGCGTGGGGTTATAAA 2404

Qy   2452 aa 2453
          ||
Db   2405 AA 2406
```

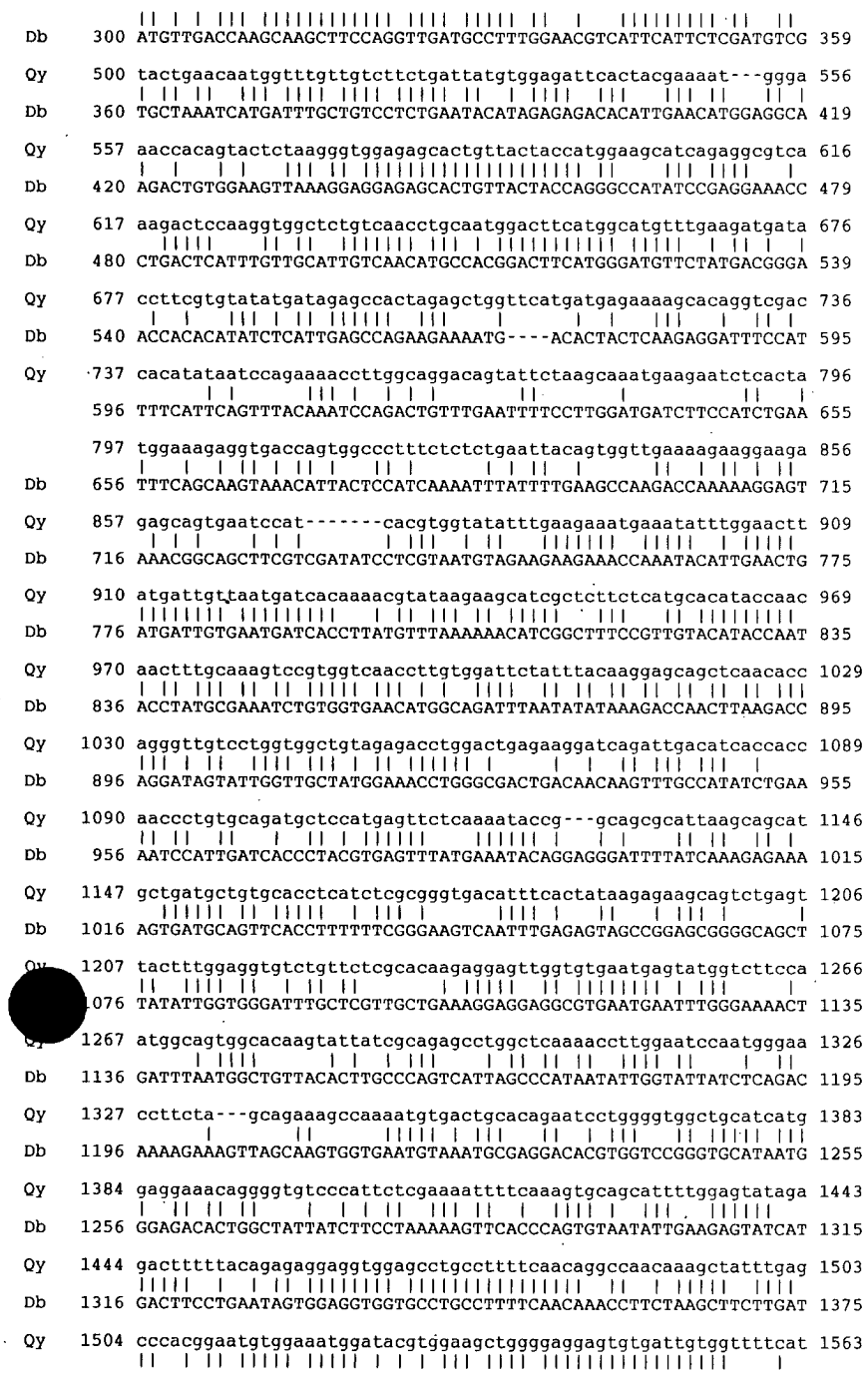
```
RESULT   7
AF155381
LOCUS       AF155381      2796 bp    mRNA            PRI       01-JUN-2000
DEFINITION  Homo sapiens metalloprotease-like, disintegrin-like, cysteine-rich
            protein 2 delta (ADAM22) mRNA, alternative splice product, complete
            cds.
ACCESSION   AF155381
VERSION     AF155381.1  GI:8132432
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2796)
  AUTHORS   Harada,T., Nishie,A., Torigoe,K., Ohtubo,M., Ikezaki,K., Shono,T.,
            Morioka,T., Wada,M. and Kuwano,M.
  TITLE     Identification of three novel splice variants of human
            metalloprotease-like, disintegrin-like, cysteine-rich protein 2
            gene and their specific expression in brain tissue and gliomas
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2796)
  AUTHORS   Wada,M., Harada,T., Nishie,A., Torigoe,K., Ohtubo,M., Nagayama,J.
            and Kuwano,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAY-1999) Biochemistry, Kyushu University Faculty of
            Medicine, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan
FEATURES             Location/Qualifiers
     source          1. .2796
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21"
     gene            1. .2796
                     /gene="ADAM22"
                     /note="MDC2"
     CDS             47. .2746
                     /gene="ADAM22"
                     /note="MDC2-delta; alternatively spliced"
                     /codon_start=1
                     /product="metalloprotease-like, disintegrin-like,
                     cysteine-rich protein 2 delta"
                     /protein_id="AAF73288.1"
                     /db_xref="GI:8132433"
                     /translation="MQAAVAVSVPFLLLCVLGTCPPARCGQAGDASLMELEKRKENRF
                     VERQSIVPLRLIYRSGGEDESRHDALDTRVRGDLGGPQLTHVDQASFQVDAFGTSFIL
                     DVVLNHDLLSSEYIERHIEHGGKTVEVKGGEHCYYQGHIRGNPDSFVALSTCHGLHGM
                     FYDGNHTYLIEPEENDTTQEDFHFHSVYKSRLFEFSLDDLPSEFQQVNITPSKFILKP
                     RPKRSKRQLRRYPRNVEEETKYIELMIVNDHLMFKKHRLSVVHTNTYAKSVVNMADLI
                     YKDQLKTRIVLVAMETWATDNKFAISENPLITLREFMKYRRDFIKEKSDAVHLFSGSQ
                     FESSRSGAAYIGGICSLLKGGGVNEFGKTDLMAVTLAQSLAHNIGIISDKRKLASGEC
                     KCEDTWSGCIMGDTGYYLPKKFTQCNIEEYHDFLNSGGGACLFNKPSKLLDPPECGNG
                     FIETGEECDCGTPAECVLEGAECCKKCTLTQDSQCSDGLCCKKCKFQPMGTVCREAVN
                     DCDIRETCSGNSSQCAPNIHKMDGYSCDGVQGICFGGRCKTRDRQCKYIWGQKVTASD
                     KYCYEKLNIEGTEKGNCGKDKDTWIQCNKRDVLCGYLLCTNIGNIPRLGELDGEITST
                     LVVQQGRTLNCSGGHVKLEEDVDLGYVEDGTPCGPQMMCLEHRCLPVASFNFSTCLSS
                     KEGTICSGNGVCSNELKCVCNRHWIGSDCNTYFPHNDDAKTGITLSGNGVAGTNIIIG
                     IIAGTILVLALILGITAWGYKNYREQRSNGLSHSWSERIPDTKHISDICENGRPRSNS
                     WQGNLGGNKKKIRGKRFRPRSNSTEYLNPWFKRDYNVAKWVEDVNKNTEEPYFRTLSP
                     AKSPSSSTGSIASSRKYPYPMPPLPDEDKKVNRQSARLWETSI"
BASE COUNT      824 a    554 c    695 g    723 t
ORIGIN
```

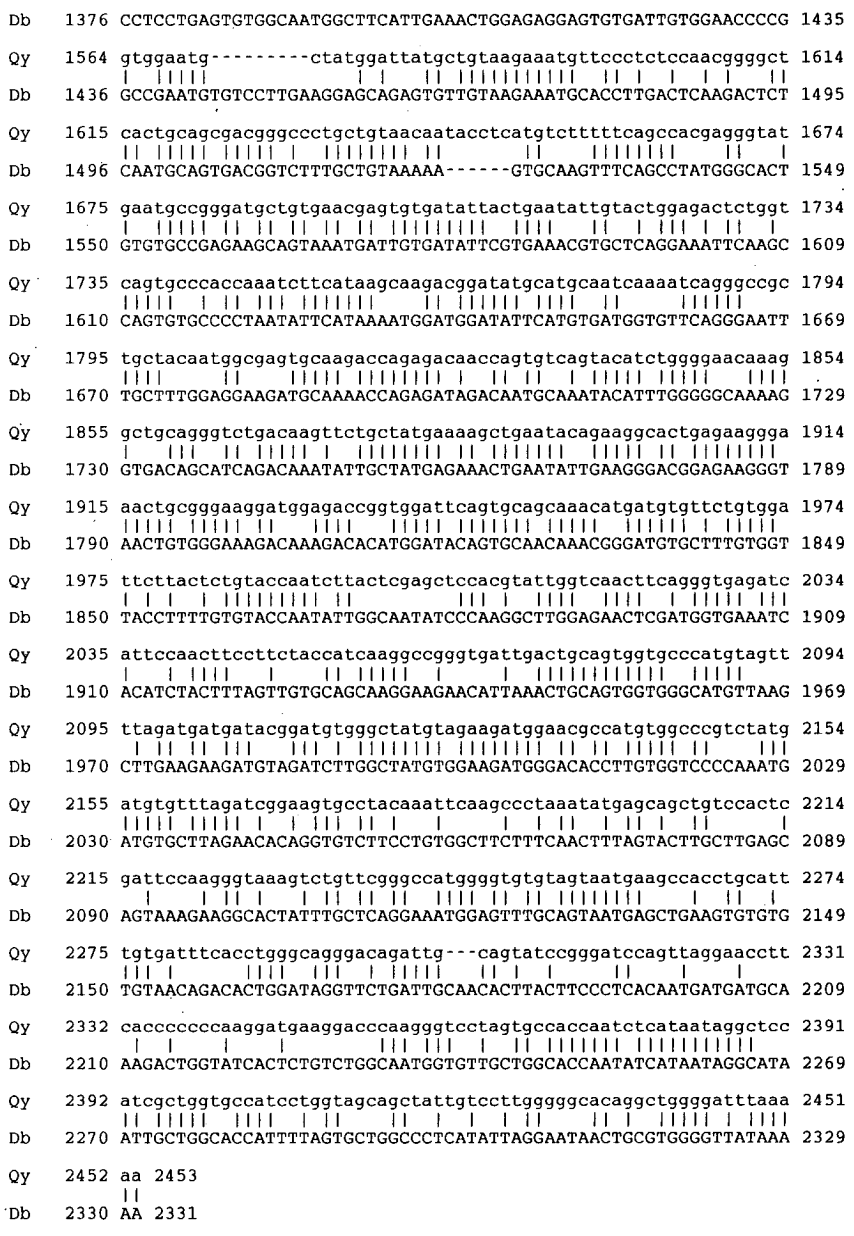
```
  Query Match           23.5%;  Score 586.8;  DB 89;  Length 2796;
  Best Local Similarity 57.2%;  Pred. No. 3.4e-130;
  Matches 1237;  Conservative    0;  Mismatches  887;  Indels   38;  Gaps    8;

Qy    320 tgcagaaagaaatcacactgccttcaagactcatatattacatcaaccaagactcggaaa 379
          || ||    | | ||   ||||     | ||||| ||   |      |   ||    ||||
Db    180 TGGAGCGCCAGAGCATCGTGCCACTGCGCCTCATCTACCGCTCGGGCGGCGAAGACGAAA 239

Qy    380 gcccttatcacgttcttgacacaaaggcaagacaccagcaaaaacataataaggctgtcc 439
          | |   |  |||  || |||||   ||   |   | | |           ||      |
Db    240 GTCGGCACGACGCGCTCGACACGCGGGTGCGGGGCGACCTCGGTGGCCCGCAGTTGACTC 299

Qy    440 atctggcccaggcaagcttccagattgaagccttcggctccaaattcattcttgacctca 499
```





RESULT 8
E26530
LOCUS E26530 2805 bp DNA PAT 07-FEB-2001
DEFINITION Novel protein belonging to MDC gene family and DNA encoding the same.
ACCESSION E26530

```
VERSION     E26530.1  GI:13026197
KEYWORDS    JP 1999155574-A/1.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2805)
  AUTHORS   Koji,S.
  TITLE     Novel protein belonging to MDC gene family and DNA encoding the
            same
  JOURNAL   Patent: JP 1999155574-A 1 15-JUN-1999;
            EISAI CO LTD
COMMENT     OS   Homo sapiens (human)
            PN   JP 1999155574-A/1
            PD   15-JUN-1999
            PF   01-DEC-1997 JP 1997330020
            PR
            PI   KOJI SONE
            PC   C12N15/09,C07K14/47,C07K16/18,C12N1/21,C12P21/02,C12Q1/68// PC
              (C12N15/09,C12R1:91),(C12N1/21,C12R1:19),(C12P21/02,C12R1:19), PC
              C12N15/00,
            PC   (C12N15/00,C12R1:91)
            CC   Strandedness: Double;
            CC   Topology: Linear;
            FH   Key             Location/Qualifiers
            FT   CDS             122. .2698.
FEATURES             Location/Qualifiers
     source          1. .2805
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      829 a    553 c    723 g    700 t
ORIGIN
```

Query Match 23.5%; Score 586.8; DB 10; Length 2805;
Best Local Similarity 57.2%; Pred. No. 3.4e-130;
Matches 1237; Conservative 0; Mismatches 887; Indels 38; Gaps 8;

```
Qy    320 tgcagaaagaaatcacactgccttcaagactcatatattacatcaacaaagactcggaaa 379
Db    255 TGGAGCGCCAGAGCATCGTGCCACTGCGCCTCATCTACCGCTCGGGCGGCGAAGACGAAA 314

Qy    380 gcccttatcacgttcttgacacaaaggcaagacaccagcaaaaacataataaggctgtcc 439
Db    315 GTCGGCACGACGCGCTCGACACGCGGGTGCGGGGCGACCTCGGTGGCCCGCAGTTGACTC 374

Qy    440 atctggcccaggcaagcttccagattgaagccttcggctccaaattcattcttgacctca 499
Db    375 ATGTTGACCAAGCAAGCTTCCAGGTTGATGCCTTTGGAACGTCATTCATTCTCGATGTCG 434

Qy    500 tactgaacaatggtttgttgtcttctgattatgtggagattcactacgaaaat---ggga 556
Db    435 TGCTAAATCATGATTTGCTGTCCTCTGAATACATAGAGAGACACATTGAACATGGAGGCA 494

Qy    557 aaccacagtactctaagggtggagagcactgttactaccatggaagcatcagaggcgtca 616
Db    495 AGACTGTGGAAGTTAAAGGAGGAGAGCACTGTTACTACCAGGGCCATATCCGAGGAAACC 554

Qy    617 aagactccaaggtggctctgtcaacctgcaatggacttcatggcatgtttgaagatgata 676
Db    555 CTGACTCATTTGTTGCATTGTCAACATGCCACGGACTTCATGGGATGTTCTATGACGGGA 614

Qy    677 ccttcgtgtatatgatagagccactagagctggttcatgatgagaaaagcacaggtcgac 736
Db    615 ACCACACATATCTCATTGAGCCAGAAGAAAATG----ACACTACTCAAGAGGATTTCCAT 670

Qy    737 cacatataatccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcacta 796
Db    671 TTTCATTCAGTTTACAAATCCAGACTGTTTGAATTTTCCTTGGATGATCTTCCATCTGAA 730

Qy    797 tggaaagaggtgaccagtggccctttctctctgaattacagtggttgaaaagaaggaaga 856
Db    731 TTTCAGCAAGTAAACATTACTCCATCAAAATTTATTTTGAAGCCAAGACCAAAAAGGAGT 790

Qy    857 gagcagtgaatccat-------cacgtggtatatttgaagaaatgaaatatttggaactt 909
Db    791 AAACGGCAGCTTCGTCGATATCCTCGTAATGTAGAAGAAGAAACCAAATACATTGAACTG 850

Qy    910 atgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgcacataccaac 969
Db    851 ATGATTGTGAATGATCACCTTATGTTTAAAAAACATCGGCTTTCCGTTGTACATACCAAT 910

Qy    970 aactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcagctcaacacc 1029
Db    911 ACCTATGCGAAATCTGTGGTGAACATGGCAGATTTAATATATAAAGACCAACTTAAGACC 970

Qy   1030 agggttgtcctggtggctgtagagacctggactgagaaggatcagattgacatcaccacc 1089
Db    971 AGGATAGTATTGGTTGCTATGGAAACCTGGGCGACTGACAACAAGTTTGCCATATCTGAA 1030

Qy   1090 aaccctgtgcagatgctccatgagttctcaaaataccg---gcagcgcattaagcagcat 1146
Db   1031 AATCCATTGATCACCCTACGTGAGTTTATGAAATACAGGAGGGATTTTATCAAAGAGAAA 1090

Qy   1147 gctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagcagtctgagt 1206
Db   1091 AGTGATGCAGTTCACCTTTTTTCGGGAAGTCAATTTGAGAGTAGCCGGAGCGGGGCAGCT 1150

Qy   1207 tactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtatggtcttcca 1266
Db   1151 TATATTGGTGGGATTTGCTCGTTGCTGAAAGGAGGAGGCGTGAATGAATTTGGGAAAACT 1210

Qy   1267 atggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatccaatgggaa 1326
Db   1211 GATTTAATGGCTGTTACACTTGCCCAGTCATTAGCCCATAATATTGGTATTATCTCAGAC 1270

Qy   1327 ccttcta---gcagaaagccaaaatgtgactgcacagaatcctggggtggctgcatcatg 1383
Db   1271 AAAAGAAAGTTAGCAAGTGGTGAATGTAAATGCGAGGACACGTGGTCCGGGTGCATAATG 1330

Qy   1384 gaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttggagtataga 1443
Db   1331 GGAGACACTGGCTATTATCTTCCTAAAAAGTTCACCCAGTGTAATATTGAAGAGTATCAT 1390

Qy   1444 gacttttttacagagaggaggtggagcctgccttttcaacaggccaacaaagctatttgag 1503
Db   1391 GACTTCCTGAATAGTGGAGGTGGTGCCTGCCTTTTCAACAAACCTTCTAAGCTTCTTGAT 1450

Qy   1504 cccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgtggttttcat 1563
Db   1451 CCTCCTGAGTGTGGCAATGGCTTCATTGAAACTGGAGAGGAGTGTGATTGTGGAACCCCG 1510

Qy   1564 gtggaatg---------ctatggattatgctgtaagaaatgttccctctccaacggggct 1614
Db   1511 GCCGAATGTGTCCTTGAAGGAGCAGAGTGTTGTAAGAAATGCACCTTGACTCAAGACTCT 1570

Qy   1615 cactgcagcgacgggccctgctgtaacaatacctcatgtctttttcagccacgagggtat 1674
Db   1571 CAATGCAGTGACGGTCTTTGCTGTAAAAA------GTGCAAGTTTCAGCCTATGGGCACT 1624

Qy   1675 gaatgccgggatgctgtgaacgagtgtgatattactgaatattgtactggagactctggt 1734
Db   1625 GTGTGCCGAGAAGCAGTAAATGATTGTGATATTCGTGAAACGTGCTCAGGAAATTCAAGC 1684

Qy   1735 cagtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatcagggccgc 1794
Db   1685 CAGTGTGCCCCTAATATTCATAAAATGGATGGATATTCATGTGATGGTGTTCAGGGAATT 1744

Qy   1795 tgctacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctggggaacaaag 1854
Db   1745 TGCTTTGGAGGAAGATGCAAAACCAGAGATAGACAATGCAAATACATTTGGGGGCAAAAG 1804

Qy   1855 gctgcagggtctgacaagttctgctatgaaaagctgaatacagaaggcactgagaaggga 1914
Db   1805 GTGACAGCATCAGACAAATATTGCTATGAGAAACTGAATATTGAAGGGACGGAGAAGGGT 1864
```

```
Qy   1915 aactgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtgga 1974
          ||||| ||||| ||   ||||   ||||| ||||||| |||||  |||||| | |||||
Db   1865 AACTGTGGGAAAGACAAAGACACATGGATACAGTGCAACAAACGGGATGTGCTTTGTGGT 1924

Qy   1975 ttcttactctgtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatc 2034
          | | |  | ||||||||| ||          ||| |  ||||  ||||  | ||||| |||
Db   1925 TACCTTTTGTGTACCAATATTGGCAATATCCCAAGGCTTGGAGAACTCGATGGTGAAATC 1984

Qy   2035 attccaacttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagtt 2094
          |   | ||||   |    ||||||||  |     |  ||||||||||||  |||||
Db   1985 ACATCTACTTTAGTTGTGCAGCAAGGAAGAACATTAAACTGCAGTGGTGGGCATGTTAAG 2044

Qy   2095 ttagatgatgatacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatg 2154
           | || || |||   ||| | |||||||| |||||||| || || ||||| ||    |||
Db   2045 CTTGAAGAAGATGTAGATCTTGGCTATGTGGAAGATGGGACACCTTGTGGTCCCCAAATG 2104

Qy   2155 atgtgtttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactc 2214
          ||||| ||||| |  | ||| || |   |     |  | ||  | || |  ||     |
Db   2105 ATGTGCTTAGAACACAGGTGTCTTCCTGTGGCTTCTTTCAACTTTAGTACTTGCTTGAGC 2164

Qy   2215 gattccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatt 2274
            |    | || |   | || || ||  |||| || || ||||||||    |  ||  |
Db   2165 AGTAAAGAAGGCACTATTTGCTCAGGAAATGGAGTTTGCAGTAATGAGCTGAAGTGTGTG 2224

Qy   2275 tgtgatttcacctgggcagggacagattg---cagtatccgggatccagttaggaacctt 2331
          ||| |      ||||  ||| | |||||   || |  |     ||    |    |
Db   2225 TGTAACAGACACTGGATAGGTTCTGATTGCAACACTTACTTCCCTCACAATGATGATGCA 2284

Qy   2332 caccccccaaggatgaaggacccaagggtcctagtgccaccaatctcataataggctcc 2391
            |  |    |   |        ||| |||  |  || |||||| ||||||||||
Db   2285 AAGACTGGTATCACTCTGTCTGGCAATGGTGTTGCTGGCACCAATATCATAATAGGCATA 2344

Qy   2392 atcgctggtgccatcctggtagcagctattgtccttgggggcacaggctggggatttaaa 2451
          || |||||  ||||  | ||    ||  |  |  | ||    || |  ||||| | ||||
Db   2345 ATTGCTGGCACCATTTTAGTGCTGGCCCTCATATTAGGAATAACTGCGTGGGGTTATAAA 2404

Qy   2452 aa 2453
          ||
Db   2405 AA 2406
```

```
RESULT   9
AF155382
LOCUS       AF155382     2858 bp    mRNA            PRI       01-JUN-2000
DEFINITION  Homo sapiens metalloprotease-like, disintegrin-like, cysteine-rich
            protein 2 epsilon (ADAM22) mRNA, complete cds.
ACCESSION   AF155382
VERSION     AF155382.1  GI:8132434
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
      NCE   1  (bases 1 to 2858)
      ORS   Harada,T., Nishie,A., Torigoe,K., Ohtubo,M., Ikezaki,K., Shono,T.,
            Morioka,T., Wada,M. and Kuwano,M.
  TITLE     Identification of three novel splice variants of human
            metalloprotease-like, disintegrin-like, cysteine-rich protein 2
            gene and their specific expression in brain tissue and gliomas
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2858)
  AUTHORS   Wada,M., Harada,T., Nishie,A., Torigoe,K., Ohtubo,M., Nagayama,J.
            and Kuwano,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-MAY-1999) Biochemistry, Kyushu University Faculty of
            Medicine, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan
FEATURES             Location/Qualifiers
     source          1. .2858
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21"
     gene            1. .2858
                     /gene="ADAM22"
                     /note="MDC2"
     CDS             47. .2767
                     /gene="ADAM22"
                     /note="MDC2-epsilon"
                     /codon_start=1
                     /product="metalloprotease-like, disintegrin-like,
                     cysteine-rich protein 2 epsilon"
                     /protein_id="AAF73289.1"
                     /db_xref="GI:8132435"
                     /translation="MQAAVAVSVPFLLLCVLGTCPPARCGQAGDASLMELEKRKENRF
                     VERQSIVPLRLIYRSGGEDESRHDALDTRVRGDLGGPQLTHVDQASFQVDAFGTSFIL
                     DVVLNHDLLSSEYIERHIEHGGKTVEVKGGEHCYYQGHIRGNPDSFVALSTCHGLHGM
                     FYDGNHTYLIEPEENDTTQEDFHFHSVYKSRLFEFSLDDLPSEFQQVNITPSKFILKP
                     RPKRSKRQLRRYPRNVEEETKYIELMIVNDHLMFKKHRLSVVHTNTYAKSVVNMADLI
                     YKDQLKTRIVLVAMETWATDNKFAISENPLITLREFMKYRRDFIKEKSDAVHLFSGSQ
                     FESSRSGAAYIGGICSLLKGGGVNEFGKTDLMAVTLAQSLAHNIGIISDKRKLASGEC
                     KCEDTWSGCIMGDTGYYLPKKFTQCNIEEYHDFLNSGGGACLFNKPSKLLDPPECGNG
                     FIETGEECDCGTPAECVLEGAECCKKCTLTQDSQCSDGLCCKKCKFQPMGTVCREAVN
                     DCDIRETCSGNSSQCAPNIHKMDGYSCDGVQGICFGGRCKTRDRQCKYIWGQKVTASD
                     KYCYEKLNIEGTEKGNCGKDKDTWIQCNKRDVLCGYLLCTNIGNIPRLGELDGEITST
                     LVVQQGRTLNCSGGHVKLEEDVDLGYVEDGTPCGPQMMCLEHRCLPVASFNFSTCLSS
                     KEGTICSGNGVCSNELKCVCNRHWIGSDCNTYFPHNDDAKTGITLSGNGVAGTNIIIG
                     IIAGTILVLALILGITAWGYKNYREQRQLPQGDYVKKPGDGDSFYSDIPPGVSTNSAS
                     SSKKRSNGLSHSWSERIPDTKHISDICENGRPRSNSWQGNLGGNKKKIRGKRFRPRSN
                     STETLSPAKSPSSSTGSIASSRKYPYPMPPLPDEDKKVNRQSARLWETSI"
BASE COUNT      835 a    578 c    708 g    737 t
ORIGIN
```

Query Match 23.5%; Score 586.8; DB 89; Length 2858;
Best Local Similarity 57.2%; Pred. No. 3.4e-130;
Matches 1237; Conservative 0; Mismatches 887; Indels 38; Gaps 8;

```
Qy    320 tgcagaaagaaatcacactgccttcaagactcatatattacatcaaccaagactcggaaa 379
          || ||    | | ||   ||||     | ||||| ||   |     |   ||    ||||
Db    180 TGGAGCGCCAGAGCATCGTGCCACTGCGCCTCATCTACCGCTCGGGCGGCGAAGACGAAA 239

Qy    380 gcccttatcacgttcttgacacaaaggcaagacaccagcaaaaacataataaggctgtcc 439
          | |   |  |||  || |||||   ||   |   | | |              ||      |
Db    240 GTCGGCACGACGCGCTCGACACGCGGGTGCGGGGCGACCTCGGTGGCCCGCAGTTGACTC 299

Qy    440 atctggcccaggcaagcttccagattgaagccttcggctccaaattcattcttgacctca 499
          || | | ||| ||||||||||||| |||| ||||| ||   |    |||||||||| ||  ||
Db    300 ATGTTGACCAAGCAAGCTTCCAGGTTGATGCCTTTGGAACGTCATTCATTCTCGATGTCG 359

Qy    500 tactgaacaatggtttgttgtcttctgattatgtggagattcactacgaaaat---ggga 556
          | || ||  ||| |||| |||| ||||| ||  | ||||  |||   ||| ||   || |
Db    360 TGCTAAATCATGATTTGCTGTCCTCTGAATACATAGAGAGACACATTGAACATGGAGGCA 419

Qy    557 aaccacagtactctaagggtggagagcactgttactaccatggaagcatcagaggcgtca 616
          |  |   | |   ||| || ||||||||||||||||||||| ||    ||| ||||   |
Db    420 AGACTGTGGAAGTTAAAGGAGGAGAGCACTGTTACTACCAGGGCCATATCCGAGGAAACC 479

Qy    617 aagactccaaggtggctctgtcaacctgcaatggacttcatggcatgtttgaagatgata 676
            |||||    || ||  ||||||| ||| | |||||||||| |||||  | || |  |
Db    480 CTGACTCATTTGTTGCATTGTCAACATGCCACGGACTTCATGGGATGTTCTATGACGGGA 539

Qy    677 ccttcgtgtatatgatagagccactagagctggttcatgatgagaaaagcacaggtcgac 736
          |  |   ||| | || ||||||  |||    |      |  |   |||   |  || |
Db    540 ACCACACATATCTCATTGAGCCAGAAGAAAATG----ACACTACTCAAGAGGATTTCCAT 595

Qy    737 cacatataatccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcacta 796
              | |      ||| |  | | |       ||       |          ||   |
Db    596 TTTCATTCAGTTTACAAATCCAGACTGTTTGAATTTTCCTTGGATGATCTTCCATCTGAA 655

Qy    797 tggaaagaggtgaccagtggccctttctctctgaattacagtggttgaaaagaaggaaga 856
          |   |  | || | || |   || |      | | ||  |      ||  | || | ||
Db    656 TTTCAGCAAGTAAACATTACTCCATCAAAATTTATTTTGAAGCCAAGACCAAAAAGGAGT 715

Qy    857 gagcagtgaatccat-------cacgtggtatatttgaagaaatgaaatatttggaactt 909
          | | |    | | |       | |||  | ||   |||||||  |||||  | |||||
Db    716 AAACGGCAGCTTCGTCGATATCCTCGTAATGTAGAAGAAGAAACCAAATACATTGAACTG 775
```

```
Qy    910 atgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgcacataccaac 969
          |||||||| |||||||||   | || ||| || |||||   |||   || ||||||||||
Db    776 ATGATTGTGAATGATCACCTTATGTTTAAAAAACATCGGCTTTCCGTTGTACATACCAAT 835

Qy    970 aactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcagctcaacacc 1029
          | || ||| || || ||||| ||| | |  ||||  || || || || || || || |||
Db    836 ACCTATGCGAAATCTGTGGTGAACATGGCAGATTTAATATATAAAGACCAACTTAAGACC 895

Qy   1030 agggttgtcctggtggctgtagagacctggactgagaaggatcagattgacatcaccacc 1089
          ||| | ||  |||| ||| | || ||||||  |     |  || ||| ||| |
Db    896 AGGATAGTATTGGTTGCTATGGAAACCTGGGCGACTGACAACAAGTTTGCCATATCTGAA 955

Qy   1090 aaccctgtgcagatgctccatgagttctcaaaataccg---gcagcgcattaagcagcat 1146
          || ||  ||   |  || | ||||||    ||||||  |   | |    || ||  || |
Db    956 AATCCATTGATCACCCTACGTGAGTTTATGAAATACAGGAGGGATTTTATCAAAGAGAAA 1015

Qy   1147 gctgatgctgtgcacctcatctcgcgggtgacatttcactataagaggaagcagtctgagt 1206
          ||||||  || |||||  | ||| |      |||| |   ||   | ||  |       |
Db   1016 AGTGATGCAGTTCACCTTTTTTCGGGAAGTCAATTTGAGAGTAGCCGGAGCGGGGCAGCT 1075

Qy   1207 tactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtatggtcttcca 1266
          ||  |||| ||  | || ||      | |||||  || ||||||| | |||     |
Db   1076 TATATTGGTGGGATTTGCTCGTTGCTGAAAGGAGGAGGCGTGAATGAATTTGGGAAAACT 1135

Qy   1267 atggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatccaatgggaa 1326
             | ||||      | |  | |||    | || || ||  |||| ||     |  ||
Db   1136 GATTTAATGGCTGTTACACTTGCCCAGTCATTAGCCCATAATATTGGTATTATCTCAGAC 1195

Qy   1327 ccttcta---gcagaaagccaaaatgtgactgcacagaatcctgggggggctgcatcatg 1383
             |      ||     ||||| | |||   ||  | |||   || |||||  |||
Db   1196 AAAAGAAAGTTAGCAAGTGGTGAATGTAAATGCGAGGACACGTGGTCCGGGTGCATAATG 1255

Qy   1384 gaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcatcttggagtataga 1443
          |  || || ||    |  | | ||  ||| ||  |  |||| |  |||   ||||||
Db   1256 GGAGACACTGGCTATTATCTTCCTAAAAAGTTCACCCAGTGTAATATTGAAGAGTATCAT 1315

Qy   1444 gacttttacagagaggaggtggagcctgccttttcaacaggccaacaaacctatttgag 1503
          |||||  |  | || |||||||| ||||||||||||||||  ||  | |||||  ||||
Db   1316 GACTTCCTGAATAGTGGAGGTGGTGCCTGCCTTTTCAACAAACCTTCCAAGCTTCTTGAT 1375

Qy   1504 cccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgtggttttcat 1563
          ||  | || ||||| ||||| | | | ||| |||| ||||||||||| |||||     |
Db   1376 CCTCCTGAGTGTGGCAATGGCTTCATTGAAACTGGAGAGGAGTGTGATTGTGGAACCCCG 1435

     1564 gtggaatg---------ctatggattatgctgtaagaaatgttccctctccaacggggct 1614
          |  |||||           | |   || ||||||||||| || |  |  | |  ||
     1436 GCCGAATGTGTCCTTGAAGGAGCAGAGTGTTGTAAGAAATGCACCTTGACTCAAGACTCT 1495

Qy   1615 cactgcagcgacgggccctgctgtaacaatacctcatgtctttttcagccacgagggtat 1674
          || ||||| ||||| |  |||||||| ||        ||    ||||||||    ||   |
Db   1496 CAATGCAGTGACGGTCTTTGCTGTAAAAA------GTGCAAGTTTCAGCCTATGGGCACT 1549

Qy   1675 gaatgccgggatgctgtgaacgagtgtgatattactgaatattgtactggagactctggt 1734
          |  ||||| || || || || || |||||||||  ||||   ||  | ||| | ||  |
Db   1550 GTGTGCCGAGAAGCAGTAAATGATTGTGATATTCGTGAAACGTGCTCAGGAAATTCAAGC 1609

Qy   1735 cagtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatcagggccgc 1794
          |||||  | || ||| |||||||    || |||||| |||  ||     ||||||
Db   1610 CAGTGTGCCCCTAATATTCATAAAATGGATGGATATTCATGTGATGGTGTTCAGGGAATT 1669

Qy   1795 tgctacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctggggaacaaag 1854
          ||||     ||    ||||| |||||||| |  || ||  | ||||| ||||   ||||
Db   1670 TGCTTTGGAGGAAGATGCAAAACCAGAGATAGACAATGCAAATACATTTGGGGGCAAAAG 1729

Qy   1855 gctgcagggtctgacaagttctgctatgaaaagctgaatacagaaggcactgagaaggga 1914
          |  |||  || ||||| |  |||||||| || |||||||  ||||| || ||||||||
Db   1730 GTGACAGCATCAGACAAATATTGCTATGAGAAACTGAATATTGAAGGGACGGAGAAGGGT 1789

Qy   1915 aactgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtgga 1974
          ||||| ||||| ||   ||||   ||||| |||||| |||||  |||||| | |||||
Db   1790 AACTGTGGGAAAGACAAAGACACATGGATACAGTGCAACAAACGGGATGTGCTTTGTGGT 1849
```

```
Qy   1975 ttcttactctgtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatc 2034
          | | |  | |||||||||| ||          ||| |  ||||  ||||  | ||||| |||
Db   1850 TACCTTTTGTGTACCAATATTGGCAATATCCCAAGGCTTGGAGAACTCGATGGTGAAATC 1909

Qy   2035 attccaacttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagtt 2094
          |   | ||||   |    || |||||  |      |  ||||||||||||  |||||
Db   1910 ACATCTACTTTAGTTGTGCAGCAAGGAAGAACATTAAACTGCAGTGGTGGGCATGTTAAG 1969

Qy   2095 ttagatgatgatacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatg 2154
            | || || |||   ||| | |||||||| |||||||| || || ||||| ||    |||
Db   1970 CTTGAAGAAGATGTAGATCTTGGCTATGTGGAAGATGGGACACCTTGTGGTCCCCAAATG 2029

Qy   2155 atgtgtttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactc 2214
          ||||| ||||| |  | ||| || | |    |    |  | ||  | || |  ||     |
Db   2030 ATGTGCTTAGAACACAGGTGTCTTCCTGTGGCTTCTTTCAACTTTAGTACTTGCTTGAGC 2089

Qy   2215 gattccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatt 2274
            |    | || |   | || || ||  |||| || || ||||||||    |  ||  |
Db   2090 AGTAAAGAAGGCACTATTTGCTCAGGAAATGGAGTTTGCAGTAATGAGCTGAAGTGTGTG 2149

Qy   2275 tgtgatttcacctgggcagggacagattg---cagtatccgggatccagttaggaacctt 2331
          ||| |      |||| |||  | |||||   || |  |     ||   |   |
Db   2150 TGTAACAGACACTGGATAGGTTCTGATTGCAACACTTACTTCCCTCACAATGATGATGCA 2209

Qy   2332 caccccccaaggatgaaggacccaagggtcctagtgccaccaatctcataataggctcc 2391
            |  |    |     |          ||| |||  |  || |||||||  ||||||||||
Db   2210 AAGACTGGTATCACTCTGTCTGGCAATGGTGTTGCTGGCACCAATATCATAATAGGCATA 2269

Qy   2392 atcgctggtgccatcctggtagcagctattgtccttgggggcacaggctggggatttaaa 2451
          || |||||  ||||  | ||    ||  |  |  | ||    || |  |||||  | ||||
Db   2270 ATTGCTGGCACCATTTTAGTGCTGGCCCTCATATTAGGAATAACTGCGTGGGGTTATAAA 2329

Qy   2452 aa 2453
          ||
Db   2330 AA 2331
```

```
RESULT  10
AF158637
LOCUS       AF158637     3259 bp    mRNA            PRI       15-SEP-1999
DEFINITION  Homo sapiens metalloproteinase-disintegrin ADAM22-3 (ADAM22) mRNA,
            alternatively spliced, partial cds.
ACCESSION   AF158637
VERSION     AF158637.1  GI:5882219
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3259)
  AUTHORS   Poindexter,K., Nelson,N., DuBose,R.F., Black,R.A. and Cerretti,D.P.
  TITLE     The identification of seven metalloproteinase-disintegrin (ADAM)
            genes from genomic libraries
  JOURNAL   Gene 237 (1), 61-70 (1999)
  MEDLINE   99453762
REFERENCE   2  (bases 1 to 3259)
  AUTHORS   Cerretti,D.P.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1999) Molecular Biology, Immunex Corporation, 51
            University St., Seattle, WA 98101, USA
FEATURES             Location/Qualifiers
     source          1. .3259
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21"
                     /tissue_type="cerebellum"
     gene            <1. .3259
                     /gene="ADAM22"
     CDS             <1. .2604
                     /gene="ADAM22"
                     /note="alternatively spliced"
                     /codon_start=1
```

```
                     /product="metalloproteinase-disintegrin ADAM22-3"
                     /protein_id="AAD55251.1"
                     /db_xref="GI:5882220"
                     /translation="KENRFVERQSIVPLRLIYRSGGEDESRHDALDTRVRGDLGGRQL
                     THVDQASFQVDAFGTSFILDVVLNHDLLSSEYIERHIEHGGKTVEVKGGEHCYYQGHI
                     RGNPDSFVALSTCHGLHGMFYDGNHTYLIEPEENDTTQEDFHFHSVYKSRLFEFSLDD
                     LPSEFQQVNITPSKFILKPRPKRSKRQLRRYPRNVEEETKYIELMIVNDHLMFKKHRL
                     SVVHTNTYAKSVVNMADLIYKDQLKTRIVLVAMETWATDNKFAISENPLITLREFMKY
                     RRDFIKEKSDAVHLFSGSQFESSRSGAAYIGGICSLLKGGGVNEFGKTDLMAVTLAQS
                     LAHNIGIISDKRKLASGECKCEDTWSGCIMGDTGYYLPKKFTQCNIEEYHDFLNSGGG
                     ACLFNKPSKLLDPPECGNGFIETGEECDCGTPAECVLEGAECCKKCTLTQDSQCSDGL
                     CCKKCKFQPMGTVCREAVNDCDIRETCSGNSSQCAPNIHKMDGYSCDGVQGICFGGRC
                     KTRDRQCKYIWGQKVTASDKYCYEKLNIEGTEKGNCGKDKDTWIQCNKRDVLCGYLLC
                     TNIGNIPRLGELDGEITSTLVVQQGRTLNCSGGHVKLEEDVDLGYVEDGTPCGPQMMC
                     LEHRCLPVASFNFSTCLSSKEGTICSGNGVCSNELKCVCNRHWIGSDCNTYFPHNDDA
                     KTGITLSGNGVAGTNIIIGIIAGTILVLALILGITAWGYKNYREQRQLPQGDYVKKPG
                     DGDSFYSDIPPGVSTNSASSSKKRSNGLSHSWSERIPDTKHISDICENGRPRSNSWQG
                     NLGGNKKKIRGKRFRPRSNSTETLSPAKSPSSSTGSIASSRKYPYPMPPLPDEDKKVN
                     RQSARLWETSI"
BASE COUNT     1009 a    617 c    738 g    888 t      7 others
ORIGIN

Query Match             23.5%;  Score 586.8;  DB 89;  Length 3259;
Best Local Similarity   57.2%;  Pred. No. 3.4e-130;
Matches 1237;  Conservative    0;  Mismatches  887;  Indels   38;  Gaps    8;

Qy     320 tgcagaaagaaatcacactgccttcaagactcatatattacatcaaccaagactcggaaa 379
Db      17 TGGAGCGCCAGAGCATCGTGCCACTGCGCCTCATCTACCGCTCGGGCGGCGAAGACGAAA 76

Qy     380 gcccttatcacgttcttgacacaaaggcaagacaccagcaaaaacataataaggctgtcc 439
Db      77 GTCGGCACGACGCGCTCGACACGCGGGTGCGGGGCGACCTCGGTGGCCGGCAGTTGACTC 136

Qy     440 atctggcccaggcaagcttccagattgaagccttcggctccaaattcattcttgacctca 499
Db     137 ATGTTGACCAAGCAAGCTTCCAGGTTGATGCCTTTGGAACGTCATTCATTCTCGATGTCG 196

Qy     500 tactgaacaatggtttgttgtcttctgattatgtggagattcactacgaaaat---ggga 556
Db     197 TGCTAAATCATGATTTGCTGTCCTCTGAATACATAGAGAGACACATTGAACATGGAGGCA 256

Qy     557 aaccacagtactctaagggtggagagcactgttactaccatggaagcatcagaggcgtca 616
Db     257 AGACTGTGGAAGTTAAAGGAGGAGAGCACTGTTACTACCAGGGCCATATCCGAGGAAACC 316

Qy     617 aagactccaaggtggctctgtcaacctgcaatggacttcatggcatgtttgaagatgata 676
Db     317 CTGACTCATTTGTTGCATTGTCAACATGCCACGGACTTCATGGGATGTTCTATGACGGGA 376

Qy     677 ccttcgtgtatatgatagagccactagagctggttcatgatgagaaaagcacaggtcgac 736
Db     377 ACCACACATATCTCATTGAGCCAGAAGAAAATG----ACACTACTCAAGAGGATTTCCAT 432

Qy     737 cacatataatccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcacta 796
Db     433 TTTCATTCAGTTTACAAATCCAGACTGTTTGAATTTTCCTTGGATGATCTTCCATCTGAA 492

Qy     797 tggaaagaggtgaccagtggccctttctctctgaattacagtggttgaaaagaaggaaga 856
Db     493 TTTCAGCAAGTAAACATTACTCCATCAAAATTTATTTTGAAGCCAAGACCAAAAAGGAGT 552

Qy     857 gagcagtgaatccat-------cacgtggtatatttgaagaaatgaaatatttggaactt 909
Db     553 AAACGGCAGCTTCGTCGATATCCTCGTAATGTAGAAGAAGAAACCAAATACATTGAACTG 612

Qy     910 atgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgcacataccaac 969
Db     613 ATGATTGTGAATGATCACCTTATGTTTAAAAAACATCGGCTTTCCGTTGTACATACCAAT 672

Qy     970 aactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcagctcaacacc 1029

Db     673 ACCTATGCGAAATCTGTGGTGAACATGGCAGATTTAATATATAAAGACCAACTTAAGACC 732

Qy    1030 agggttgtcctggtggctgtagagacctggactgagaaggatcagattgacatcaccacc 1089
Db     733 AGGATAGTATTGGTTGCTATGGAAACCTGGGCGACTGACAACAAGTTTGCCATATCTGAA 792

Qy    1090 aaccctgtgcagatgctccatgagttctcaaaataccg---gcagcgcattaagcagcat 1146
Db     793 AATCCATTGATCACCCTACGTGAGTTTATGAAATACAGGAGGGATTTTATCAAAGAGAAA 852

Qy    1147 gctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagcagtctgagt 1206
Db     853 AGTGATGCAGTTCACCTTTTTTCGGGAAGTCAATTTGAGAGTAGCCGGAGCGGGGCAGCT 912

Qy    1207 tactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtatggtcttcca 1266
Db     913 TATATTGGTGGGATTTGCTCGTTGCTGAAAGGAGGAGGCGTGAATGAATTTGGGAAAACT 972

Qy    1267 atggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatccaatgggaa 1326
Db     973 GATTTAATGGCTGTTACACTTGCCCAGTCATTAGCCCATAATATTGGTATTATCTCAGAC 1032

Qy    1327 ccttcta---gcagaaagccaaaatgtgactgcacagaatcctggggtggctgcatcatg 1383
Db    1033 AAAAGAAAGTTAGCAAGTGGTGAATGTAAATGCGAGGACACGTGGTCCGGGTGCATAATG 1092

Qy    1384 gaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttggagtataga 1443
Db    1093 GGAGACACTGGCTATTATCTTCCTAAAAAGTTCACCCAGTGTAATATTGAAGAGTATCAT 1152

Qy    1444 gacttttacagagaggaggtggagcctgccttttcaacaggccaacaaagctatttgag 1503
Db    1153 GACTTCCTGAATAGTGGAGGTGGTGCCTGCCTTTTCAACAAACCTTCTAAGCTTCTTGAT 1212

Qy    1504 cccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgtggttttcat 1563
Db    1213 CCTCCTGAGTGTGGCAATGGCTTCATTGAAACTGGAGAGGAGTGTGATTGTGGAACCCCG 1272

Qy    1564 gtggaatg---------ctatggattatgctgtaagaaatgttccctctccaacggggct 1614
Db    1273 GCCGAATGTGTCCTTGAAGGAGCAGAGTGTTGTAAGAAATGCACCTTGACTCAAGACTCT 1332

Qy    1615 cactgcagcgacgggccctgctgtaacaatacctcatgtctttttcagccacgagggtat 1674
Db    1333 CAATGCAGTGACGGTCTTTGCTGTAAAAA------GTGCAAGTTTCAGCCTATGGGCACT 1386

Qy    1675 gaatgccgggatgctgtgaacgagtgtgatattactgaatattgtactggagactctggt 1734
Db    1387 GTGTGCCGAGAAGCAGTAAATGATTGTGATATTCGTGAAACGTGCTCAGGAAATTCAAGC 1446

Qy    1735 cagtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatcagggccgc 1794
Db    1447 CAGTGTGCCCCTAATATTCATAAAATGGATGGATATTCATGTGATGGTGTTCAGGGAATT 1506

Qy    1795 tgctacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctggggaacaaag 1854
Db    1507 TGCTTTGGAGGAAGATGCAAAACCAGAGATAGACAATGCAAATACATTTGGGGGCAAAAG 1566

Qy    1855 gctgcagggtctgacaagttctgctatgaaaagctgaatacagaaggcactgagaaggga 1914
Db    1567 GTGACAGCATCAGACAAATATTGCTATGAGAAACTGAATATTGAAGGGACGGAGAAGGGT 1626

Qy    1915 aactgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtgga 1974
Db    1627 AACTGTGGGAAAGACAAAGACACATGGATACAGTGCAACAAACGGGATGTGCTTTGTGGT 1686

Qy    1975 ttcttactctgtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatc 2034
Db    1687 TACCTTTTGTGTACCAATATTGGCAATATCCCAAGGCTTGGAGAACTCGATGGTGAAATC 1746

Qy    2035 attccaacttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagtt 2094
Db    1747 ACATCTACTTTAGTTGTGCAGCAAGGAAGAACATTAAACTGCAGTGGTGGGCATGTTAAG 1806
```

```
Qy   2095 ttagatgatgatacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatg 2154
Db   1807 CTTGAAGAAGATGTAGATCTTGGCTATGTGGAAGATGGGACACCTTGTGGTCCCCAAATG 1866

Qy   2155 atgtgtttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactc 2214
Db   1867 ATGTGCTTAGAACACAGGTGTCTTCCTGTGGCTTCTTTCAACTTTAGTACTTGCTTGAGC 1926

Qy   2215 gattccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatt 2274
Db   1927 AGTAAAGAAGGCACTATTTGCTCAGGAAATGGAGTTTGCAGTAATGAGCTGAAGTGTGTG 1986

Qy   2275 tgtgatttcacctgggcagggacagattg---cagtatccgggatccagttaggaacctt 2331
Db   1987 TGTAACAGACACTGGATAGGTTCTGATTGCAACACTTACTTCCCTCACAATGATGATGCA 2046

Qy   2332 cacccccccaaggatgaaggacccaagggtcctagtgccaccaatctcataataggctcc 2391
Db   2047 AAGACTGGTATCACTCTGTCTGGCAATGGTGTTGCTGGCACCAATATCATAATAGGCATA 2106

Qy   2392 atcgctggtgccatcctggtagcagctattgtccttgggggcacaggctggggatttaaa 2451
Db   2107 ATTGCTGGCACCATTTTAGTGCTGGCCCTCATATTAGGAATAACTGCGTGGGGTTATAAA 2166

Qy   2452 aa 2453
Db   2167 AA 2168
```

```
RESULT  11
AF073291
LOCUS       AF073291     3295 bp    mRNA            PRI       04-JAN-2000
DEFINITION  Homo sapiens MDC2 (MDC2) mRNA, complete cds.
ACCESSION   AF073291
VERSION     AF073291.2  GI:6670842
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3295)
  AUTHORS   Wada,M., Torigoe,K., Harada,T. and Kuwano,M.
  TITLE     Isolation and tissue specific expression of novel ADAM family from
            7q21.1 region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 3295)
  AUTHORS   Wada,M., Torigoe,K., Harada,T. and Kuwano,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-1998) Biochemistry, Kyushu University School of
            Medicine, Maidashi 1-1, Higasi-ku, Fukuoka, Fukuoka 812-8582, Japan
REFERENCE   3  (bases 1 to 3295)
  AUTHORS   Wada,M., Torigoe,K., Harada,T. and Kuwano,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JAN-2000) Biochemistry, Kyushu University School of
            Medicine, Maidashi 1-1, Higasi-ku, Fukuoka, Fukuoka 812-8582, Japan
  REMARK    Sequence update by submitter
COMMENT     On Jan 4, 2000 this sequence version replaced gi:6652795.
FEATURES             Location/Qualifiers
     source          1. .3295
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q21.1"
                     /tissue_type="brain"
     gene            1. .3295
                     /gene="MDC2"
     CDS             47. .2659
                     /gene="MDC2"
                     /codon_start=1
                     /product="MDC2"
                     /protein_id="AAF22476.2"
                     /db_xref="GI:6670843"
                     /translation="MQAAVAVSVPFLLLCVLGTCPPARCGQAGDASLMELEKRKENRF
                     VERQSIVPLRLIYRSGGEDESRHDALDTRVRGDLGGRQLTHVDQASFQVDAFGTSFIL
                     DVVLNHDLLSSEYIERHIEHGGKTVEVKGGEHCYYQGHIRGNPDSFVALSTCHGLHGM
                     FYDGNHTYLIEPEENDTTQEDFHFHSVYKSRLFEFSLDDLPSEFQQVNITPSKFILKP
                     RPKRSKRQLRRYPRNVEEETKYIELMIVNDHLMFKKHRLSVVHTNTYAKSVVNMADLI
                     YKDQLKTRIVLVAMETWATDNKFAISENPLITLREFMKYRRDFIKEKSDAVHLFSGSQ
                     FESSRSGAAYIGGICSLLKGGGVNEFGKTDLMAVTLAQSLAHNIGIISDKRKLASGEC
                     KCEDTWSGCIMGDTGYYLPKKFTQCNIEEYHDFLNSGGGACLFNKPSKLLDPPECGNG
                     FIETGEECDCGTPAECVLEGAECCKKCTLTQDSQCSDGLCCKKCKFQPMGTVCREAVN
                     DCDIRETCSGNSSQCAPNIHKMDGYSCDGVQGICFGGRCKTRDRQCKYIWGQKVTASD
                     KYCYEKLNIEGTEKGNCGKDKDTWIQCNKRDVLCGYLLCTNIGNIPRLGELDGEITST
                     LVVQQGRTLNCSGGHVKLEEDVDLGYVEDGTPCGPQMMCLEHRCLPVASFNFSTCLSS
                     KEGTICSGNGVCSNELKCVCNRHWIGSDCNTYFPHNDDAKTGITLSGNGVAGTNIIIG
                     IIAGTILVLALILGITAWGYKNYREQRSNGLSHSWSERIPDTKHISDICENGRPRSNS
                     WQGNLGGNKKKIRGKRFRPRSNSTETLSPAKSPSSSTGSIASSRKYPYPMPPLPDEDK
                     KVNRQSARLWETSI"
BASE COUNT      985 a    641 c    779 g    887 t      3 others
ORIGIN
```

```
Query Match             23.5%;  Score 586.8;  DB 88;  Length 3295;
Best Local Similarity   57.2%;  Pred. No. 3.4e-130;
Matches 1237;  Conservative    0;  Mismatches  887;  Indels   38;  Gaps    8;

Qy    320 tgcagaaagaaatcacactgccttcaagactcatatattacatcaaccaagactcggaaa 379
Db    180 TGGAGCGCCAGAGCATCGTGCCACTGCGCCTCATCTACCGCTCGGGCGGCGAAGACGAAA 239

Qy    380 gcccttatcacgttcttgacacaaaggcaagacaccagcaaaaacataataaggctgtcc 439
Db    240 GTCGGCACGACGCGCTCGACACGCGGGTGCGGGGCGACCTCGGTGGCCGGCAGTTGACTC 299

Qy    440 atctggcccaggcaagcttccagattgaagccttcggctccaaattcattcttgacctca 499
Db    300 ATGTTGACCAAGCAAGCTTCCAGGTTGATGCCTTTGGAACGTCATTCATTCTCGATGTCG 359

Qy    500 tactgaacaatggtttgttgtcttctgattatgtggagattcactacgaaaat---ggga 556
Db    360 TGCTAAATCATGATTTGCTGTCCTCTGAATACATAGAGAGACACATTGAACATGGAGGCA 419

Qy    557 aaccacagtactctaagggtggagagcactgttactaccatggaagcatcagaggcgtca 616
Db    420 AGACTGTGGAAGTTAAAGGAGGAGAGCACTGTTACTACCAGGGCCATATCCGAGGAAACC 479

Qy    617 aagactccaaggtggctctgtcaacctgcaatggacttcatggcatgtttgaagatgata 676
Db    480 CTGACTCATTTGTTGCATTGTCAACATGCCACGGACTTCATGGGATGTTCTATGACGGGA 539

Qy    677 ccttcgtgtatatgatagagccactagagctggttcatgatgagaaaagcacaggtcgac 736
Db    540 ACCACACATATCTCATTGAGCCAGAAGAAAATG----ACACTACTCAAGAGGATTTCCAT 595

Qy    737 cacatataatccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcacta 796
Db    596 TTTCATTCAGTTTACAAATCCAGACTGTTTGAATTTTCCTTGGATGATCTTCCATCTGAA 655

Qy    797 tggaaagaggtgaccagtggccctttctctctgaattacagtggttgaaaagaaggaaga 856
Db    656 TTTCAGCAAGTAAACATTACTCCATCAAAATTTATTTTGAAGCCAAGACCAAAAAGGAGT 715

Qy    857 gagcagtgaatccat-------cacgtggtatatttgaagaaatgaaatatttggaactt 909
Db    716 AAACGGCAGCTTCGTCGATATCCTCGTAATGTAGAAGAAGAAACCAAATACATTGAACTG 775

Qy    910 atgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgcacataccaac 969
Db    776 ATGATTGTGAATGATCACCTTATGTTTAAAAAACATCGGCTTTCCGTTGTACATACCAAT 835

Qy    970 aactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcagctcaacacc 1029
Db    836 ACCTATGCGAAATCTGTGGTGAACATGGCAGATTTAATATATAAAGACCAACTTAAGACC 895

Qy   1030 agggttgtcctggtggctgtagagacctggactgagaaggatcagattgacatcaccacc 1089
```

```
Db    896 AGGATAGTATTGGTTGCTATGGAAACCTGGGCGACTGACAACAAGTTTGCCATATCTGAA 955

Qy   1090 aaccctgtgcagatgctccatgagttctcaaaataccg---gcagcgcattaagcagcat 1146
Db    956 AATCCATTGATCACCCTACGTGAGTTTATGAAATACAGGAGGGATTTTATCAAAGAGAAA 1015

Qy   1147 gctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagcagtctgagt 1206
Db   1016 AGTGATGCAGTTCACCTTTTTTCGGGAAGTCAATTTGAGAGTAGCCGGAGCGGGGCAGCT 1075

Qy   1207 tactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtatggtcttcca 1266
Db   1076 TATATTGGTGGGATTTGCTCGTTGCTGAAAGGAGGAGGCGTGAATGAATTTGGGAAAACT 1135

Qy   1267 atggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatccaatgggaa 1326
Db   1136 GATTTAATGGCTGTTACACTTGCCCAGTCATTAGCCCATAATATTGGTATTATCTCAGAC 1195

Qy   1327 ccttcta---gcagaaagccaaaatgtgactgcacagaatcctggggtggctgcatcatg 1383
     1196 AAAAGAAAGTTAGCAAGTGGTGAATGTAAATGCGAGGACACGTGGTCCGGGTGCATAATG 1255

     1384 gaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttggagtataga 1443
Db   1256 GGAGACACTGGCTATTATCTTCCTAAAAAGTTCACCCAGTGTAATATTGAAGAGTATCAT 1315

Qy   1444 gacttttacagagaggaggtggagcctgccttttcaacaggccaacaaagctatttgag 1503
Db   1316 GACTTCCTGAATAGTGGAGGTGGTGCCTGCCTTTTCAACAAACCTTCTAAGCTTCTTGAT 1375

Qy   1504 cccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgtggttttcat 1563
Db   1376 CCTCCTGAGTGTGGCAATGGCTTCATTGAAACTGGAGAGGAGTGTGATTGTGGAACCCCG 1435

Qy   1564 gtggaatg---------ctatggattatgctgtaagaaatgttccctctccaacggggct 1614
Db   1436 GCCGAATGTGTCCTTGAAGGAGCAGAGTGTTGTAAGAAATGCACCTTGACTCAAGACTCT 1495

Qy   1615 cactgcagcgacgggccctgctgtaacaatacctcatgtctttttcagccacgagggtat 1674
Db   1496 CAATGCAGTGACGGTCTTTGCTGTAAAAA------GTGCAAGTTTCAGCCTATGGGCACT 1549

Qy   1675 gaatgccgggatgctgtgaacgagtgtgatattactgaatattgtactggagactctggt 1734
Db   1550 GTGTGCCGAGAAGCAGTAAATGATTGTGATATTCGTGAAACGTGCTCAGGAAATTCAAGC 1609

Qy   1735 cagtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatcagggccgc 1794
Db   1610 CAGTGTGCCCCTAATATTCATAAAATGGATGGATATTCATGTGATGGTGTTCAGGGAATT 1669

Qy   1795 tgctacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctggggaacaaag 1854
     1670 TGCTTTGGAGGAAGATGCAAAACCAGAGATAGACAATGCAAATACATTTGGGGGCAAAAG 1729

Qy   1855 gctgcagggtctgacaagttctgctatgaaaagctgaatacagaaggcactgagaaggga 1914
Db   1730 GTGACAGCATCAGACAAATATTGCTATGAGAAACTGAATATTGAAGGGACGGAGAAGGGT 1789

Qy   1915 aactgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtgga 1974
Db   1790 AACTGTGGGAAAGACAAAGACACATGGATACAGTGCAACAAACGGGATGTGCTTTGTGGT 1849

Qy   1975 ttcttactctgtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatc 2034
Db   1850 TACCTTTTGTGTACCAATATTGGCAATATCCCAAGGCTTGGAGAACTCGATGGTGAAATC 1909

Qy   2035 attccaacttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagtt 2094
Db   1910 ACATCTACTTTAGTTGTGCAGCAAGGAAGAACATTAAACTGCAGTGGTGGGCATGTTAAG 1969

Qy   2095 ttagatgatgatacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatg 2154
Db   1970 CTTGAAGAAGATGTAGATCTTGGCTATGTGGAAGATGGGACACCTTGTGGTCCCCAAATG 2029

Qy   2155 atgtgtttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactc 2214
Db   2030 ATGTGCTTAGAACACAGGTGTCTTCCTGTGGCTTCTTTCAACTTTAGTACTTGCTTGAGC 2089

Qy   2215 gattccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatt 2274
Db   2090 AGTAAAGAAGGCACTATTTGCTCAGGAAATGGAGTTTGCAGTAATGAGCTGAAGTGTGTG 2149

Qy   2275 tgtgatttcacctgggcagggacagattg---cagtatccgggatccagttaggaacctt 2331
Db   2150 TGTAACAGACACTGGATAGGTTCTGATTGCAACACTTACTTCCCTCACAATGATGATGCA 2209

Qy   2332 caccccccaaggatgaaggacccaagggtcctagtgccaccaatctcataataggctcc 2391
Db   2210 AAGACTGGTATCACTCTGTCTGGCAATGGTGTTGCTGGCACCAATATCATAATAGGCATA 2269

Qy   2392 atcgctggtgccatcctggtagcagctattgtccttgggggcacaggctggggatttaaa 2451
Db   2270 ATTGCTGGCACCATTTTAGTGCTGGCCCTCATATTAGGAATAACTGCGTGGGGTTATAAA 2329

Qy   2452 aa 2453
Db   2330 AA 2331
```

```
RESULT  12
AB009674
LOCUS       AB009674     2773 bp    mRNA            ROD       17-AUG-1999
DEFINITION  Mus musculus mRNA for ADAM22, complete cds.
ACCESSION   AB009674
VERSION     AB009674.1  GI:5736625
KEYWORDS    ADAM22.
SOURCE      Mus musculus brain cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Sagane,K., Yamazaki,K., Mizui,Y. and Tanaka,I.
  TITLE     Cloning and chromosomal mapping of mouse ADAM11, ADAM22 and ADAM23
  JOURNAL   Gene 236 (1), 79-86 (1999)
  MEDLINE   99365303
REFERENCE   2  (bases 1 to 2773)
  AUTHORS   Sagane,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-1997) to the DDBJ/EMBL/GenBank databases. Koji
            Sagane, Eisai Co.,Ltd., Tsukuba Research Laboratories; Tokodai
            5-1-3, Tsukuba, Ibaraki 300-2635, Japan
            (E-mail:k-sagane@hhc.eisai.co.jp, Tel:+81-298-47-5813,
            Fax:+81-298-47-5367)
FEATURES             Location/Qualifiers
     source          1. .2773
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /tissue_type="brain"
     CDS             join(124. .2417,2526. .2697)
                     /codon_start=1
                     /product="ADAM22 beta"
                     /protein_id="BAA83383.1"
                     /db_xref="GI:5736627"
                     /translation="MQAAAAASFWLLCVLGTCPLARCGRAGVASLKGLERGKENRFLE
                     RQSIIPLRLIYRLGGEDETQHNQLDTRVRGDPGGPQLTHVDKASFRVDAFGTSFVLDV
                     LLNHELLSSGYVERQIEHGGKVVENKGGEHCYYQGQIRGNPVSFVALSTCHGLHGMFY
                     DGNHTYLIEPEENEKSQESSHCHSVYKSRQFEFPLDDLPSEFQRVNITPPQFILKPRL
                     KRRKRQLLRFPRNVEEETKYIELMIVNDHLMFKKHRLSVVYTNTYAKSVVNMADVIYK
                     DQLKTRIVLVAMETWAADNKFAISENPLITLREFMKYRRDFIKEKADAVHLFSGSQFE
                     SSRSGAAYIGGICSLLRGGGVNEFGKTDLMAVTLAQSLAHNVGIISDKRKLASGECKC
                     EDTWSGCIMGDTGYYLPKKFTQCNVEEYHDFLNSGGGACLFNKPSKLLDPPECGNGFI
                     ETGEECDCGTPAECALEGAECCKKCTLTQDSQCSDGLCCKKCKFQPLGTVCREAVNDC
                     DIREICSGNSSQCAPNVHKMDGYSCDGTQGICFGGRCKTRDRQCKYIWGQKVTASDRY
                     CYEKLNIEGTEKGNCGKDKDTWTQCNKRDVLCGYLLCTNIGNIPRLGELDGEITSTLV
                     VQQGRTLNCSGAHVKLEEDVDLGYVEDGTPCGPQMMCLEHRCLPVASFNFSTCSSSKA
```

```
                     GTVCSGNGVCSNELKCVCNRHWTGADCGTHFPHNDDAKTGITLSGNGVAGTNIIIGII
                     AGTILVLALILGITAWGYKNYREQRSNGLSHSWSERIPDTKHISDICENGRPRSNSWQ
                     GNMGGNKKKIRGKRFRPRSNSTE"
     CDS             124. .2697
                     /codon_start=1
                     /product="ADAM22 alpha"
                     /protein_id="BAA83382.1"
                     /db_xref="GI:5736626"
                     /translation="MQAAAAASFWLLCVLGTCPLARCGEAGVASLKGLERGKENRFLE
                     RQSIIPLRLIYRLGGEDETQHNQLDTRVRGDPGGPQLTEVDKASFRVDAFGTSFVLDV
                     LLNHELLSSGYVERQIEHGGKVVENKGGEHCYYQGQIRGNPVSFVALSTCHGLHGMFY
                     DGNHTYLIEPEENEKSQESSHCHSVYKSRQFEFPLDDLFSEFQRVNITPPQFILKPRL
                     KRRKRQLLRFPRNVEEETKYIELMIVNDHLMFKKHRLSVVYTNTYAKSVVNMADVIYK
                     DQLKTRIVLVAMETWAADNKFAISENPLITLREFMEYRRDFIKEKADAVHLFSGSQFE
                     SSRSGAAYIGGICSLLRGGGVNEFGKTDLMAVTLAQSLAHNVGIISDKRKLASGECKC
                     EDTWSGCIMGDTGYYLPKKFTQCNVEEYHDFLNSGGGACLFNKPSKLLDPPECGNGFI
                     ETGEECDCGTPAECALEGAECCKKCTLTQDSQCSDGLCCKKCKFQPLGTVCREAVNDC
                     DIREICSGNSSQCAPNVHKMDGYSCDGTQGICFGGHCKTRDRQCKYIWGQKVTASDRY
                     CYEKLNIEGTEKGNCGKDKDTWTQCNKRDVLCGYLLCTNIGNIPRLGELDGEITSTLV
                     VQQGRTLNCSGAHVKLEEDVDLGYVEDGTPCGPQMMCLEHRCLPVASFNFSTCSSSKA
                     GTVCSGNGVCSNELKCVCNRHWTGADCGTHFPHNDDAKTGITLSGNGVAGTNIIIGII
                     AGTILVLALILGITAWGYKNYREQRQLPQGDYVKKPGDGDSFYSDFPPGGSTNSASSS
                     KKRSNGLSHSWSERIPDTKHISDICENGRPRSNSWQGNMGGNKKKIRGKRFRPRSNST
                     E"
     polyA_site      2773
                     /note="10 a nucleotides"
BASE COUNT      765 a    610 c    757 g    641 t
ORIGIN
```

```
  Query Match             23.1%;  Score 576.8;  DB 94;  Length 2773;
  Best Local Similarity   56.6%;  Pred. No. 8.6e-128;
  Matches 1221;  Conservative    0;  Mismatches  907;  Indels   30;  Gaps    7;

Qy     320 tgcagaaagaaatcacactgccttcaagactcatatattacatcaccaagactcggaaa 379
           || ||    | | ||   | ||     | ||||| ||| | |   |   ||    ||||
Db     251 TGGAGCGCCAGAGCATCATACCACTGCGCCTCATCTATCGCTTGGGCGGCGAGGATGAAA 310

Qy     380 gcccttatcacgttcttgacacaaaggcaagacaccagcaaaaacataataaggctgtcc 439
             |   ||   || |||||    ||    |   | | |           ||      |
Db     311 CTCAGCACAACCAGCTCGACACGCGGGTGCGGGGCGACCCTGGCGGACCGCAGCTGACGC 370

Qy     440 atctggcccaggcaagcttccagattgaagccttcggctccaaattcattcttgacctca 499
           || ||| | ||||||||||||| | | || ||||| ||  ||   ||| |||| ||| ||
Db     371 ATGTGGACAAGGCAAGCTTCCGGGTGGACGCCTTTGGAACCTCTTTCGTTCTCGACGTCC 430

       500 tactgaacaatggtttgttgtcttctgattatgtggagattcactacgaaaat---ggga 556
           | |||||| | |  ||| |||| || | |||||| |||| ||    ||| ||   || |
       431 TGCTGAACCACGAGTTGCTGTCGTCAGGTTATGTAGAGAGACAGATTGAACATGGAGGCA 490

Qy     557 aaccacagtactctaagggtggagagcactgttactaccatggaagcatcagaggcgtca 616
           |      | |    || || ||||| |||||||| ||||| ||    ||  ||||   |
Db     491 AGGTCGTGGAGAACAAAGGAGGAGAACACTGTTATTACCAGGGCCAGATTCGAGGAAACC 550

Qy     617 aagactccaaggtggctctgtcaacctgcaatggacttcatggcatgtttgaagatgata 676
             | |||    || || || ||||| ||| ||||||| ||||| |||||  | ||||  |
Db     551 CTGTCTCATTTGTTGCCCTATCAACATGCCATGGACTCCATGGGATGTTCTATGATGGAA 610

Qy     677 ccttcgtgtatatgatagagccactagagctggttcatgatgagaaaagcacaggtcgac 736
             |  |   ||| | || ||||||  |||    |   |  | |  |     |   | | |
Db     611 ACCACACATATCTCATTGAGCCAGAAGAAAATGAGAAGTCTCAAGAGTCGTCTCATTGTC 670

Qy     737 cacatataatccag---aaaaccttggcaggacagtattctaagcaaatgaagaatctca 793
               | |   | |   || | |   |  |   | | |    |   |   |  |
Db     671 ACTCCGTGTACAAATCCAGACAGTTTGAATTTCCCTTGGATGACCTTCCGTCTGAATTTC 730

Qy     794 ctatggaaagaggtgaccagtggccctttctctctgaattacagtggttgaaaagaagga 853
             | ||    |   |    |  ||| | |   |   |        ||  ||||  |
Db     731 AACGCGTAAACATTACTCCCCCTCAGTTTATTTTGAAGCCAAGACTAAAGAGGAGAAAAA 790

Qy     854 agagagcagtgaatccatcacgtggtatatttgaagaaatgaaatatttggaacttatga 913
                 |        | ||   | ||   || ||||  |||||  | ||||| ||||
Db     791 GACAGCTTCTTCGATTTCCTCGGAATGTAGAGGAGGAAACCAAATACATCGAACTGATGA 850
```

```
Qy     914 ttgttaatgatcacaaaacgtataagaagcatcgctcttctcatgcacataccaacaact 973
           |||| ||||||||    | || ||| ||||| ||   ||||  || | |||| || | ||
Db     851 TTGTGAATGATCATCTCATGTTTAAAAAGCACCGGCTTTCTGTTGTATATACTAATACCT 910

Qy     974 ttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcagctcaacaccaggg 1033
            ||| ||||| |||||  ||| | |  |||   || || || || || || || || |||
Db     911 ATGCGAAGTCTGTGGTGAACATGGCAGATGTGATCTATAAAGACCAACTGAAAACTAGGA 970

Qy     1034 ttgtcctggtggctgtagagacctggactgagaaggatcagattgacatcaccaccaacc 1093
            ||||  |||| ||  | || |||||| |||   |  |  || ||| |||  |    ||||
Db     971 TTGTATTGGTTGCCATGGAAACCTGGGCTGCTGACAACAAGTTTGCCATATCTGAAAACC 1030

Qy     1094 ctgtgcagatgctccatgagttctcaaaataccg---gcagcgcattaagcagcatgctg 1150
            |  ||   |  || | |||||||   |||||| |   | |    || ||| || | ||||
Db     1031 CACTGATTACCCTTCGTGAGTTCATGAAATACAGGAGGGATTTTATCAAGGAGAAGGCTG 1090

Qy     1151 atgctgtgcacctcatctcgcgggtgacatttcactataagagaagcagtctgagttact 1210
            |||| || || ||  ||||| |      |||| |   ||   | ||| |      |||
Db     1091 ATGCAGTTCATCTTTTCTCGGGCAGTCAATTTGAGAGTAGCCGGAGCGGGGCAGCTTATA 1150

Qy     1211 ttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtatggtcttccaatgg 1270
            |||| ||  | || || |      |||||||  || |||||||| | |||     |
Db     1151 TTGGTGGGATTTGCTCGCTGCTGAGAGGAGGAGGCGTGAATGAATTTGGAAAAACGGATT 1210

Qy     1271 cagtggcacaagtattatcgcagagcctggctcaaaaccttggaatccaatgggaacctt 1330
            | |||||       |  | |||   || || || ||  | || ||    |  ||
Db     1211 TAATGGCAGTTACCCTCGCTCAGTCACTAGCCCACAATGTCGGTATAATCTCTGACAAAA 1270

Qy     1331 ctagcag---aaagccaaaatgtgactgcacagaatcctggggtggctgcatcatggagg 1387
              |       ||      | ||| | ||    ||  |||||   || ||||| |||  |
Db     1271 GAAAACTAGCAAGTGGTGAGTGTAAATGTGAGGACACCTGGTCCGGGTGCATAATGGGAG 1330

Qy     1388 aaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttggagtatagagact 1447
            | || ||    |  |   |  ||| ||  | ||||||   |   ||| |||   ||||
Db     1331 ACACCGGCTACTATCTCCCGAAAAAGTTCACTCAGTGCAATGTCGAGGAATATCATGACT 1390

Qy     1448 ttttacagagaggaggtggagcctgccttttcaacaggccaacaaagctatttgagccca 1507
            |  |  | || || || || |||||||| ||||||| |||  | |||||  |||| ||
Db     1391 TCCTGAATAGTGGTGGAGGTGCCTGCCTGTTCAACAAGCCTTCCAAGCTTCTTGATCCTC 1450

Qy     1508 cggaatgtggaaatggatacgtggaagctggggaggagtgtgattgtggttttcatgtgg 1567
            | || |||||||| || | | | ||| | || ||||||||||||||||    | ||  |
Db     1451 CTGAGTGTGGAAACGGCTTCATTGAAACCGGAGAGGAGTGTGATTGTGGCACCCCTGCAG 1510

Qy     1568 aatg---------ctatggattatgctgtaagaaatgttcccctctccaacggggctcact 1618
            ||||           | |   || |||||||||||  || |  || | |   |||| |
Db     1511 AATGTGCCCTTGAAGGGGCAGAGTGTTGTAAGAAATGCACCTTGACCCAGGATTCTCAAT 1570

Qy     1619 gcagcgacgggccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaat 1678
            |||| || || |  || ||||| ||        ||    |||||||| |  |||   | |
Db     1571 GCAGTGATGGCCTTTGTTGTAAAAA------GTGCAAGTTTCAGCCTCTGGGGACCGTGT 1624

Qy     1679 gccgggatgctgtgaacgagtgtgatattactgaatattgtactggagactctggtcagt 1738
            |||| || || |||||||| ||||| |||  ||||   ||| | ||| | ||  | ||||
Db     1625 GCCGAGAGGCAGTGAACGACTGTGACATTCGTGAAATATGTTCAGGAAATTCGAGCCAGT 1684

Qy     1739 gcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgct 1798
            |  | || ||| | |||||    || |||||| | ||  ||   | ||||||    ||||
Db     1685 GTGCCCCCAATGTGCATAAAATGGATGGATATTCGTGTGATGGTACTCAGGGAATTTGCT 1744

Qy     1799 acaatggcgagtgcaagaccagagacaaccagtgtcagtacatctggggaacaaaggctg 1858
               ||    || || ||||||||| |  || ||  | ||||||||||   |||||
Db     1745 TTGGAGGAAGATGTAAAACCAGAGATAGACAATGCAAATACATCTGGGGGCAAAAGGTAA 1804

Qy     1859 cagggtctgacaagttctgctatgaaaagctgaatacagaaggcactgagaagggaaact 1918
            |||  |||||| || |||||||| || ||||| |  || || || |||||||| ||||
Db     1805 CAGCATCTGACAGGTACTGCTATGAGAAACTGAACATTGAGGGGACGGAGAAGGGTAACT 1864

Qy     1919 gcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtggattct 1978
            | ||||| ||    |||  |||||  ||||||| |||||  |||||  | ||||| | |
Db     1865 GTGGGAAAGACAAGGACACGTGGACACAGTGCAACAAACGGGATGTACTTTGTGGTTACC 1924
```

```
Qy   1979 tactctgtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatcattc 2038
          | || || || ||  ||          ||| |  | ||  | ||| | ||||||||||
Db   1925 TCCTGTGCACAAACATTGGCAACATCCCAAGGCTCGGGGAGCTTGATGGTGAGATCACAT 1984

Qy   2039 caacttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagttttag 2098
          | || |   |    || |||||  |      |  |||||||||| || |||||     | |
Db   1985 CCACATTAGTTGTGCAGCAAGGAAGAACATTAAACTGCAGTGGCGCGCATGTTAAACTTG 2044

Qy   2099 atgatgatacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatgatgt 2158
          | || |||  |||| | |||||||| |||||||| ||||| |||||||    |||||||
Db   2045 AAGAAGATGTGGATCTTGGCTATGTGGAAGATGGGACGCCCTGTGGCCCTCAGATGATGT 2104

Qy   2159 gtttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactcgatt 2218
          | || || |  | ||| || |   |     || | ||  | || |  ||  |   |  |
Db   2105 GCTTGGAACACAGGTGTCTTCCTGTGGCTTCCTTCAACTTCAGTACTTGCTCCAGCAGTA 2164

Qy   2219 ccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatttgtg 2278
               || |  || || || ||  |||| || || || |||||   ||  ||  | ||
Db   2165 AAGCAGGCACCGTGTGCTCAGGAAATGGAGTTTGCAGCAATGAGCTCAAGTGTGTGTGCA 2224

Qy   2279 atttcacctgggcagggacagattgcagtatccgggatccagttaggaaccttt---cacc 2335
          |      |||| ||||| | || ||| | | ||    |||    |   |   |    |
Db   2225 ACAGGCACTGGACAGGGGCTGACTGCGGCACCCACTTTCCTCACAATGATGATGCAAAGA 2284

Qy   2336 cccccaaggatgaaggacccaagggtcctagtgccaccaatctcataataggctccatcg 2395
          |   ||           ||| |||  |  || ||| ||| ||||||||||    || |
Db   2285 CTGGCATCACCCTGTCTGGCAATGGTGTTGCTGGCACTAATATCATAATAGGAATAATTG 2344

Qy   2396 ctggtgccatcctggtagcagctattgtccttgggggcacaggctggggatttaaaaa 2453
          | ||  ||||  | ||    ||  |  || | ||    || | |||||| | ||||||
Db   2345 CCGGCACCATTTTAGTGCTGGCCCTCATCTTAGGAATAACTGCCTGGGGTTATAAAAA 2402


RESULT  13
AB009675
LOCUS       AB009675     3168 bp    mRNA            PRI       15-AUG-1998
DEFINITION  Homo sapiens mRNA for MDC/ADAM11, complete cds.
ACCESSION   AB009675
VERSION     AB009675.1  GI:3419879
KEYWORDS    MDC/ADAM11.
SOURCE      Homo sapiens brain cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3168)
  AUTHORS   Sagane,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-1997) to the DDBJ/EMBL/GenBank databases. Koji
            Sagane, Eisai Co.,Ltd., Tsukuba Research Laboratories; Tokodai
            5-1-3, Tsukuba, Ibaraki 300-2635, Japan
            (E-mail:kl-sagane@eisai.co.jp, Tel:+81-298-47-5813,
            Fax:+81-298-47-5367)
      NCE   2  (sites)
      ORS   Sagane,K., Ohya,Y., Hasegawa,Y. and Tanaka,I.
      E     Metalloproteinase-like, disintegrin-like, cysteine-rich proteins
            MDC2 and MDC3: novel human cellular disintegrins highly expressed
            in the brain
  JOURNAL   Biochem. J. 334 (Pt 1), 93-98 (1998)
  MEDLINE   98359734
COMMENT     Sequence updated (24-Jul-1998).
FEATURES             Location/Qualifiers
     source          1. .3168
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_type="brain"
     CDS             1. .2310
                     /codon_start=1
                     /product="MDC/ADAM11"
                     /protein_id="BAA32352.1"
                     /db_xref="GI:3419880"
                     /translation="MRLLRRWAFAALLLSLLPTPGLGTQGPAGALRWGGLPQLGGPGA
                     PEVTEPSRLVRESSGGEVRKQQLDTRVRQEPPGGPPVHLAQVSFVIPAFNSNFTLDLE
                     LNHHLLSSQYVERHFSREGTTQHSTGAGDHCYYQGKLRGNPHSFAALSTCQGLHGVFS
                     DGNLTYIVEPQEVAGPWGAPQGPLPHLIYRTPLLPDPLGCREPGCLFAVPAQSAPPNR
                     PRLRRKRQVRRGHPTVHSETKYVELIVINDHQLFEQMRQSVVLTSNFAKSVVNLADVI
                     YKEQLNTRIVLVAMETWADGDKIQVQDDLLETLARLMVYRREGLPEPSDATHLFSGRT
                     FQSTSSGAAYVGGICSLSHGGGVNEYGNMGAMAVTLAQTLGQNLGMMWNKHRSSAGDC
                     KCPDIWLGCIMEDTGFYLPRKFSRCSIDEYNQFLQEGGGSCLFNKPLKLLDPPECGNG
                     FVEAGEECDCGSVQECSRAGGNCCKKCTLTHDAMCSDGLCCRRCKYEPRGVSCREAVN
                     ECDIAETCTGDSSQCPPNLHKLDGYYCDHEQGRCYGGRCKTRDRQCQVLWGHAAADRF
                     CYEKLNVEGTERGSCGRKGSGWVQCSKQDVLCGFLLCVNISGAPRLGDLVGDISSVTF
                     YHQGKELDCRGGHVQLADGSDLSYVEDGTACGPNMLCLDHRCLPASAFNFSTCPGSGE
                     RRICSHHGVCSNEGKCICQPDWTGKDCSIHNPLPTSPPTGETERYKGPSGTNIIIGSI
                     AGAVLVAAIVLGGTGWGFKNIRRGRSGGA"
     polyA_site      3168
                     /note="15 a nucleotides"
BASE COUNT      629 a    982 c    980 g    577 t
ORIGIN


  Query Match             23.1%;  Score 576.6;  DB 85;  Length 3168;
  Best Local Similarity   57.6%;  Pred. No. 9.5e-128;
  Matches 1227;  Conservative    0;  Mismatches  859;  Indeis   45;  Gaps    9;

Qy    376 gaaagcccttatcacgttcttgacacaaaggcaagacaccagcaaaaacataataaggct 435
          ||   ||    | ||    || |||||||  ||   | ||  ||| |  |        | ||
Db    181 GAGGTCCGAAAGCAGCAGCTGGACACAAGGGTCCGCCAGGAGCCACCAGGGGGCCCGCCT 240

Qy    436 gtccatctggcccaggcaagcttccagattgaagccttcggctccaaattcattcttgac 495
          ||||||||||||||||  || |||   ||   |||||||  ||| || ||||  || |||
Db    241 GTCCATCTGGCCCAGGTGAGTTTCGTCATCCCAGCCTTCAACTCAAACTTCACCCTGGAC 300

Qy    496 ctcatactgaacaatggtttgttgtcttctgattatgtggagattcactac------gaa 549
          ||    |||||| |      |  | || ||  | || ||||||    |||| |      ||
Db    301 CTGGAGCTGAACCACCACCTCCTCTCCTCGCAATACGTGGAGCGCCACTTCAGCCGGGAG 360

Qy    550 aatgggaaaccacagtactctaagggtggagagcactgttactaccatggaagcatcaga 609
              |  |  ||   | |   || |||||| ||||| |||||||| || |   || |
Db    361 GGGACAACCCAGCACAGCACCGGGGCTGGAGACCACTGCTACTACCAGGGGAAGCTCCGG 420

Qy    610 ggcgtcaaagactccaaggtggctctgtcaacctgcaatggacttcatggcatgtttgaa 669
          ||   |    |||||   |  || || || |||||| | || || |||||  | ||
Db    421 GGGAACCCGCACTCCTTCGCCGCCCTCTCCACCTGCCAGGGGCTGCATGGGGTCTTCTCT 480

Qy    670 gatgataccttcgtgtatatgatagagccactagagctggttcatgatgagaaaagcac- 728
          ||||  | |||    || ||  | ||||| | |||| ||| |     |  |  | | |
Db    481 GATGGGAACTTGACTTACATCGTGGAGCCCCAAGAGGTGGCTGGACCTTGGGGAGCCCCT 540

Qy    729 -----aggtcgaccacatataatccagaaaaccttggcaggacagtattctaagcaaatg 783
               |  | || ||  | ||  |    |||           |   || |      |
Db    541 CAGGGACCCCTTCCCCACCTCATTTACCGGACCCCTCTCCTCCCAGATCCCCTCGGATGC 600

Qy    784 aagaatctcactatggaaagaggtgaccagtggccctttctctctgaattacagtggttg 843
          | | | |               | ||  |     |  |      ||  |  | |  | |
Db    601 AGGGAACCAGGCTGCCTGTTTGCTGTGCCTGCCCAGTCGGCTCCTCCAAACCGGCCGAGG 660

Qy    844 aaaagaaggaagagagcagtgaatccatcacgtggtatatttga---agaaatgaaatat 900
            |||||||| ||    ||     |     |   || | | |    ||||  || |||
Db    661 CTGAGAAGGAAAAGGCAGGTCCGCCGGGGCCACCCTACAGTGCACAGTGAAACCAAGTAT 720

Qy    901 ttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgca 960
            |||| || ||  |  | || || ||| |   ||   || ||   ||    ||    |
Db    721 GTGGAGCTAATTGTGATCAACGACCACCAGCTGTTCGAGCAGATGCGACAGTCGGTGGTC 780

Qy    961 cataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcag 1020
          |  |||| |||||||| |||||||||| ||||| |  |||   || ||||||||||||
Db    781 CTCACCAGCAACTTTGCCAAGTCCGTGGTGAACCTGGCCGATGTGATATACAAGGAGCAG 840

Qy   1021 ctcaacaccagggttgtcctggtggctgtagagacctggactgagaaggatcagattgac 1080
          ||||||||  |  | |||||||| ||  | || || ||| | ||   |||  ||||  |
Db    841 CTCAACACTCGCATCGTCCTGGTTGCCATGGAAACATGGGCAGATGGGGACAAGATCCAG 900

Qy   1081 atcaccaccaaccctgtgcagatgctccatgagttctcaaaatac---cggcagcgcatt 1137
          |         |||   || ||| ||       | ||      |||   ||| || |  |
Db    901 GTGCAGGATGACCTCCTGGAGACCCTGGCCCGGCTCATGGTCTACCGACGGGAGGGTCTG 960
```

```
Qy   1138 aagcagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagc 1197
Db    961 CCTGAGCCCAGTGATGCCACCCACCTCTTCTCGGGCAGGACCTTCCAGAGCACGAGCAGC 1020

Qy   1198 agtctgagttactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtat 1257
Db   1021 GGGGCAGCCTACGTGGGGGGCATATGCTCCCTGTCCCATGGCGGGGGTGTGAACGAGTAC 1080

Qy   1258 ggtcttccaatggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatc 1317
Db   1081 GGCAACATGGGGGCGATGGCCGTGACCCTTGCCCAGACGCTGGGACAGAACCTGGGCATG 1140

Qy   1318 caatgggaaccttctagcagaaagccaa---aatgtgactgcacagaatcctggggtggc 1374
     1141 ATGTGGAACAAACACCGGAGCTCGGCAGGGGACTGCAAGTGTCCAGACATCTGGCTGGGC 1200

     1375 tgcatcatggaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttg 1434
Db   1201 TGCATCATGGAGGACACTGGGTTCTACCTGCCCCGCAAGTTCTCTCGCTGCAGCATCGAC 1260

Qy   1435 gagtatagagacttttacagagaggaggtggagcctgccttttcaacaggccaacaaag 1494
Db   1261 GAGTACAACCAGTTTCTGCAGGAGGGTGGTGGCAGCTGCCTCTTCAACAAGCCCCTCAAG 1320

Qy   1495 ctatttgagcccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgt 1554
Db   1321 CTCCTGGACCCCCCAGAGTGCGGGAACGGCTTCGTGGAGGCAGGGGAGGAGTGCGACTGC 1380

Qy   1555 ggttttcatgtggaatgc---------tatggattatgctgtaagaaatgttccctctcc 1605
Db   1381 GGCTCGGTGCAGGAGTGCAGCCGCGCAGGTGGCAACTGCTGCAAGAAATGCACCCTGACT 1440

Qy   1606 aacggggctcactgcagcgacgggccctgctgtaacaatacctcatgtctttttcagcca 1665
Db   1441 CACGACGCCATGTGCAGCGACGGGCTCTGCTG------TCGCCGCTGCAAGTACGAACCA 1494

Qy   1666 cgagggtatgaatgccgggatgctgtgaacgagtgtgatattactgaatattgtactgga 1725
Db   1495 CGGGGTGTGTCCTGCCGAGAGGCCGTGAACGAGTGCGACATCGCGGAGACCTGCACCGGG 1554

Qy   1726 gactctggtcagtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaat 1785
Db   1555 GACTCTAGCCAGTGCCCGCCTAACCTGCACAAGCTGGACGGTTACTACTGTGACCATGAG 1614

Qy   1786 cagggccgctgctacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctgg 1845
Db   1615 CAGGGCCGCTGCTACGGAGGTCGCTGCAAAACCCGGGACCGGCAGTGCCAGGTTCTTTGG 1674

     1846 ggaacaaaggctgcagggtctgacaagttctgctatgaaaagctgaatacagaaggcact 1905
Db   1675 GGCCATGCGGCTG------CTGATCGCTTCTGCTACGAGAAGCTGAATGTGGAGGGGACG 1728

Qy   1906 gagaagggaaactgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtg 1965
Db   1729 GAGCGTGGGAGCTGTGGGCGCAAGGGATCCGGCTGGGTCCAGTGCAGTAAGCAGGACGTG 1788

Qy   1966 ttctgtggattcttactctgtaccaatcttactcgagctccacgtattggtcaacttcag 2025
Db   1789 CTGTGTGGCTTCCTCCTCTGTGTCAACATCTCTGGAGCTCCTCGGCTAGGGGACCTGGTG 1848

Qy   2026 ggtgagatcattccaacttccttctaccatcaaggccgggtgattgactgcagtggtgcc 2085
Db   1849 GGAGACATCAGTAGTGTCACCTTCTACCACCAGGGCAAGGAGCTGGACTGCAGGGGAGGC 1908

Qy   2086 catgtagttttagatgatgatacggatgtgggctatgtagaagatggaacgccatgtggc 2145
Db   1909 CACGTGCAGCTGGCGGACGGCTCTGACCTGAGCTATGTGGAGGATGGCACAGCCTGCGGG 1968

Qy   2146 ccgtctatgatgtgtttagatcggaagtgcctacaaattcaagccctaaatatgagcagc 2205
Db   1969 CCTAACATGTTGTGCCTGGACCATCGCTGCCTGCCAGCTTCTGCCTTCAACTTCAGCACC 2028

Qy   2206 tgtccactcgattccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagcc 2265
```

```
Db   2029 TGCCCCGGCAGTGGGGAGCGCCGGATTTGCTCCCACCACGGGGTCTGCAGCAATGAAGGG 2088

Qy   2266 acctgcatttgtgatttcacctgggcagggacagattgcagtatccgggatccagttagg 2325
Db   2089 AAGTGCATCTGTCAGCCAGACTGGACAGGCAAAGACTGCAGTATCCATAACCCCCTGCCC 2148

Qy   2326 aaccttcacccccccaaggat---gaaggacccaagggtcctagtgccaccaatctcata 2382
Db   2149 ACGTCCCCACCCACGGGGGAGACGGAGAGATATAAAGGTCCCAGCGGCACCAACATCATC 2208

Qy   2383 ataggctccatcgctggtgccatcctggtagcagctattgtccttgggggcacaggctgg 2442
Db   2209 ATTGGCTCCATCGCTGGGGCTGTCCTGGTTGCAGCCATCGTCCTGGGCGGCACGGGCTGG 2268

Qy   2443 ggatttaaaaatgtcaagaagagaaggttcg 2473
Db   2269 GGATTTAAAAACATTCGCCGAGGAAGGTCCG 2299


RESULT  14
I25848
LOCUS       I25848       3183 bp    DNA             PAT       07-OCT-1996
DEFINITION  Sequence 8 from patent US 5552526.
ACCESSION   I25848
VERSION     I25848.1  GI:1605718
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3183)
  AUTHORS   Nakamura,Y. and Emi,M.
  TITLE     MDC proteins and DNAS encoding the same
  JOURNAL   Patent: US 5552526-A 8 03-SEP-1996;
FEATURES             Location/Qualifiers
     source          1. .3183
                     /organism="unknown"
BASE COUNT      645 a    982 c    979 g    577 t
ORIGIN


  Query Match           23.0%;  Score 575;  DB 10;  Length 3183;
  Best Local Similarity   57.5%;  Pred. No. 2.3e-127;
  Matches 1226;  Conservative    0;  Mismatches  860;  Indels   45;  Gaps    9;

Qy    376 gaaagcccttatcacgttcttgacacaaaggcaagacaccagcaaaaacataataaggct 435
Db    181 GAGGTCCGAAAGCAGCAGCTGGACACAAGGGTCCGCCAGGAGCCACCAGGGGGCCCGCCT 240

Qy    436 gtccatctggcccaggcaagcttccagattgaagccttcggctccaaattcattcttgac 495
Db    241 GTCCATCTGGCCCAGGTGAGTTTCGTCATCCCAGCCTTCAACTCAAACTTCACCCTGGAC 300

Qy    496 ctcatactgaacaatggtttgttgtcttctgattatgtggagattcactac------gaa 549
Db    301 CTGGAGCTGAACCACCACCTCCTCTCCTCGCAATACGTGGAGCGCCACTTCAGCCGGGAG 360

Qy    550 aatgggaaaccacagtactctaagggtggagagcactgttactaccatggaagcatcaga 609
Db    361 GGGACAACCCAGCACAGCACCGGGGCTGGAGACCACTGCTACTACCAGGGGAAGCTCCGG 420

Qy    610 ggcgtcaaagactccaaggtggctctgtcaacctgcaatggacttcatggcatgtttgaa 669
Db    421 GGGAACCCGCACTCCTTCGCCGCCCTCTCCACCTGCCAGGGGCTGCATGGGGTCTTCTCT 480

Qy    670 gatgataccttcgtgtatatgatagagccactagagctggttcatgatgagaaaagcac- 728
Db    481 GATGGGAACTTGACTTACATCGTGGAGCCCCAAGAGGTGGCTGGACCTTGGGGAGCCCCT 540

Qy    729 -----aggtcgaccacatataatccagaaaaccttggcaggacagtattctaagcaaatg 783
Db    541 CAGGGACCCCTTCCCCACCTCATTTACCGGACCCCTCTCCTCCCAGATCCCCTCGGATGC 600
```

```
Qy    784 aagaatctcactatggaaagaggtgaccagtggccctttctctctgaattacagtggttg 843
Db    601 AGGGAACCAGGCTGCCTGTTTGCTGTGCCTGCCCAGTCGGCTCCTCCAAACCGGCCGAGG 660

Qy    844 aaaagaaggaagagagcagtgaatccatcacgtggtatatttga---agaaatgaaaatat 900
Db    661 CTGAGAAGGAAAAGGCAGGTCCGCCGGGGCCACCCTACAGTGCACAGTGAAACCAAGTAT 720

Qy    901 ttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgca 960
Db    721 GTGGAGCTAATTGTGATCAACGACCACCAGCTGTTCGAGCAGATGCGACAGTCGGTGGTC 780

Qy    961 cataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcag 1020
Db    781 CTCACCAGCAACTTTGCCAAGTCCGTGGTGAACCTGGCCGATGTGATATACAAGGAGCAG 840

Qy   1021 ctcaacaccagggttgtcctggtggctgtagagacctggactgagaaggatcagattgac 1080
Db    841 CTCAACACTCGCATCGTCCTGGTTGCCATGGAAACATGGGCAGATGGGGACAAGATCCAG 900

Qy   1081 atcaccaccaaccctgtgcagatgctccatgagttctcaaaatac---cggcagcgcatt 1137
Db    901 GTGCAGGATGACCTCCTGGAGACCCTGGCCCGGCTCATGGTCTACCGACGGGAGGGTCTG 960

Qy   1138 aagcagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagc 1197
Db    961 CCTGAGCCCAGTAATGCCACCCACCTCTTCTCGGGCAGGACCTTCCAGAGCACGAGCAGC 1020

Qy   1198 agtctgagttactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtat 1257
Db   1021 GGGGCAGCCTACGTGGGGGGCATATGCTCCCTGTCCCATGGCGGGGGTGTGAACGAGTAC 1080

Qy   1258 ggtcttccaatggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatc 1317
Db   1081 GGCAACATGGGGGCGATGGCCGTGACCCTTGCCCAGACGCTGGGACAGAACCTGGGCATG 1140

Qy   1318 caatgggaaccttctagcagaaagccaa---aatgtgactgcacagaatcctggggtggc 1374
Db   1141 ATGTGGAACAAACACCGGAGCTCGGCAGGGGACTGCAAGTGTCCAGACATCTGGCTGGGC 1200

Qy   1375 tgcatcatggaggaaacaggggtgtcccattctcgaaaatttttcaaagtgcagcattttg 1434
Db   1201 TGCATCATGGAGGACACTGGGTTCTACCTGCCCCGCAAGTTCTCTCGCTGCAGCATCGAC 1260

Qy   1435 gagtatagagacttttttacagagaggaggtggagcctgccttttcaacaggccaacaaag 1494
Db   1261 GAGTACAACCAGTTTCTGCAGGAGGGTGGTGGCAGCTGCCTCTTCAACAAGCCCCTCAAG 1320

Qy   1495 ctatttgagcccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgt 1554
Db   1321 CTCCTGGACCCCCCAGAGTGCGGGAACGGCTTCGTGGAGGCAGGGGAGGAGTGCGACTGC 1380

Qy   1555 ggttttcatgtggaatgc---------tatggattatgctgtaagaaatgttccctctcc 1605
Db   1381 GGCTCGGTGCAGGAGTGCAGCCGCGCAGGTGGCAACTGCTGCAAGAAATGCACCCTGACT 1440

Qy   1606 aacggggctcactgcagcgacgggccctgctgtaacaatacctcatgtctttttcagcca 1665
Db   1441 CACGACGCCATGTGCAGCGACGGGCTCTGCTG------TCGCCGCTGCAAGTACGAACCA 1494

Qy   1666 cgagggtatgaatgccgggatgctgtgaacgagtgtgatattactgaatattgtactgga 1725
Db   1495 CGGGGTGTGTCCTGCCGAGAGGCCGTGAACGAGTGCGACATCGCGGAGACCTGCACCGGG 1554

Qy   1726 gactctggtcagtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaat 1785
Db   1555 GACTCTAGCCAGTGCCCGCCTAACCTGCACAAGCTGGACGGTTACTACTGTGACCATGAG 1614

Qy   1786 cagggccgctgctacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctgg 1845
Db   1615 CAGGGCCGCTGCTACGGAGGTCGCTGCAAAACCCGGGACCGGCAGTGCCAGGTTCTTTGG 1674

Qy   1846 ggaacaaaggctgcagggtctgacaagttctgctatgaaaagctgaatacagaaggcact 1905
```

```
Db   1675 GGCCATGCGGCTG------CTGATCGCTTCTGCTACGAGAAGCTGAATGTGGAGGGGACG 1728

Qy   1906 gagaagggaaactgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtg 1965
Db   1729 GAGCGTGGGAGCTGTGGGCGCAAGGGATCCGGCTGGGTCCAGTGCAGTAAGCAGGACGTG 1788

Qy   1966 ttctgtggattcttactctgtaccaatcttactcgagctccacgtattggtcaacttcag 2025
Db   1789 CTGTGTGGCTTCCTCCTCTGTGTCAACATCTCTGGAGCTCCTCGGCTAGGGGACCTGGTG 1848

Qy   2026 ggtgagatcattccaacttccttctaccatcaaggccgggtgattgactgcagtggtgcc 2085
Db   1849 GGAGACATCAGTAGTGTCACCTTCTACCACCAGGGCAAGGAGCTGGACTGCAGGGGAGGC 1908

Qy   2086 catgtagttttagatgatgatacggatgtgggctatgtagaagatggaacgccatgtggc 2145
Db   1909 CACGTGCAGCTGGCGGACGGCTCTGACCTGAGCTATGTGGAGGATGGCACAGCCTGCGGG 1968

Qy   2146 ccgtctatgatgtgtttagatcggaagtgcctacaaattcaagccctaaatatgagcagc 2205
Db   1969 CCTAACATGTTGTGCCTGGACCATCGCTGCCTGCCAGCTTCTGCCTTCAACTTCAGCACC 2028

Qy   2206 tgtccactcgattccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagcc 2265
Db   2029 TGCCCCGGCAGTGGGGAGCGCCGGATTTGCTCCCACCACGGGGTCTGCAGCAATGAAGGG 2088

Qy   2266 acctgcatttgtgatttcacctgggcagggacagattgcagtatccgggatccagttagg 2325
Db   2089 AAGTGCATCTGTCAGCCAGACTGGACAGGCAAAGACTGCAGTATCCATAACCCCCTGCCC 2148

Qy   2326 aaccttcacccccccaaggat---gaaggacccaagggtcctagtgccaccaatctcata 2382
Db   2149 ACGTCCCCACCCACGGGGGAGACGGAGAGATATAAAGGTCCCAGCGGCACCAACATCATC 2208

Qy   2383 ataggctccatcgctggtgccatcctggtagcagctattgtccttgggggcacaggctgg 2442
Db   2209 ATTGGCTCCATCGCTGGGGCTGTCCTGGTTGCAGCCATCGTCCTGGGCGGCACGGGCTGG 2268

Qy   2443 ggatttaaaaatgtcaagaagagaaggttcg 2473
Db   2269 GGATTTAAAAACATTCGCCGAGGAAGGTCCG 2299
```

```
RESULT  15
I43485
LOCUS       I43485       3183 bp    DNA             PAT       07-OCT-1997
DEFINITION  Sequence 8 from patent US 5631351.
ACCESSION   I43485
VERSION     I43485.1  GI:2468729
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 3183)
  AUTHORS   Nakamura,Y. and Emi,M.
  TITLE     Antibodies to MDC proteins
  JOURNAL   Patent: US 5631351-A 8 20-MAY-1997;
FEATURES             Location/Qualifiers
     source          1. .3183
                     /organism="unknown"
BASE COUNT      645 a    982 c    979 g    577 t
ORIGIN
```

```
  Query Match           23.0%;  Score 575;  DB 10;  Length 3183;
  Best Local Similarity   57.5%;  Pred. No. 2.3e-127;
  Matches 1226;  Conservative    0;  Mismatches  860;  Indels   45;  Gaps    9;

Qy    376 gaaagcccttatcacgttcttgacacaaaggcaagacaccagcaaaaacataataaggct 435
Db    181 GAGGTCCGAAAGCAGCAGCTGGACACAAGGGTCCGCCAGGAGCCACCAGGGGGCCCGCCT 240
```

```
Qy    436 gtccatctggcccaggcaagcttccagattgaagccttcggctccaaattcattcttgac 495
          ||||||||||||||||  || |||   ||   |||||||  ||| || ||||  || |||
Db    241 GTCCATCTGGCCCAGGTGAGTTTCGTCATCCCAGCCTTCAACTCAAACTTCACCCTGGAC 300

Qy    496 ctcatactgaacaatggtttgttgtcttctgattatgtggagattcactac------gaa 549
          ||    |||||| |      |  | || ||  | || ||||||   |||| |       ||
Db    301 CTGGAGCTGAACCACCACCTCCTCTCCTCGCAATACGTGGAGCGCCACTTCAGCCGGGAG 360

Qy    550 aatgggaaaccacagtactctaagggtggagagcactgttactaccatggaagcatcaga 609
              |  ||   | |   || |||||| |||| |||||||| || |    || |
Db    361 GGGACAACCCAGCACAGCACCGGGGCTGGAGACCACTGCTACTACCAGGGGAAGCTCCGG 420

Qy    610 ggcgtcaaagactccaaggtggctctgtcaacctgcaatggacttcatggcatgtttgaa 669
          ||   |    |||||    |  || || || |||||| | || || |||||   | ||
      421 GGGAACCCGCACTCCTTCGCCGCCCTCTCCACCTGCCAGGGGCTGCATGGGGTCTTCTCT 480

      670 gatgataccttcgtgtatatgatagagccactagagctggttcatgatgagaaaagcac- 728
          ||||  | |||    || ||  | ||||| | |||| ||| |     |  |  |  | |
Db    481 GATGGGAACTTGACTTACATCGTGGAGCCCCAAGAGGTGGCTGGACCTTGGGGAGCCCCT 540

Qy    729 -----aggtcgaccacatataatccagaaaaccttggcaggacagtattctaagcaaatg 783
              |   | || ||  | ||  |    |||          |  || |       |
Db    541 CAGGGACCCCTTCCCCACCTCATTTACCGGACCCCTCTCCTCCCAGATCCCCTCGGATGC 600

Qy    784 aagaatctcactatggaaagaggtgaccagtggccctttctctctgaattacagtggttg 843
          | | | |          | ||  |    |  |       ||  |   | |  |  |
Db    601 AGGGAACCAGGCTGCCTGTTTGCTGTGCCTGCCCAGTCGGCTCCTCCAAACCGGCCGAGG 660

Qy    844 aaaagaaggaagagagcagtgaatccatcacgtggtatatttga---agaaatgaaatat 900
          ||||||||  ||    ||     |    || | |  |    ||||  || |||
Db    661 CTGAGAAGGAAAAGGCAGGTCCGCCGGGGCCACCCTACAGTGCACACGTGAAACCAAGTAT 720

Qy    901 ttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgca 960
          |||| || ||  |  | || || ||| |   ||   || ||   ||    ||    |
Db    721 GTGGAGCTAATTGTGATCAACGACCACCAGCTGTTCGAGCAGATGCCACAGTCGGTGGTC 780

Qy    961 cataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcag 1020
          |  |||| |||||||||| |||||||||| ||||| |  |||   || ||||||||||||
Db    781 CTCACCAGCAACTTTGCCAAGTCCGTGGTGAACCTGGCCGATGTGATATACAAGGAGCAG 840

Qy   1021 ctcaacaccagggttgtcctggtggctgtagagacctggactgagaaggatcagattgac 1080
          ||||||||  |  | ||||||||  ||  | || || ||| | ||   |||  ||||  |
Db    841 CTCAACACTCGCATCGTCCTGGTTGCCATGGAAACATGGGCAGATGGGGACAAGATCCAG 900

Qy   1081 atcaccaccaaccctgtgcagatgctccatgagttctcaaaatac---cggcagcgcatt 1137
          |     |||   || |||  ||      | ||      |||   |||  || |  |
Db    901 GTGCAGGATGACCTCCTGGAGACCCTGGCCCGGCTCATGGTCTACCGACCGGAGGGTCTG 960

Qy   1138 aagcagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagc 1197
               |||    | ||||   |||||| ||||| |   |||  || ||    | ||| |||
Db    961 CCTGAGCCCAGTAATGCCACCCACCTCTTCTCGGGCAGGACCTTCCAGACCACGAGCAGC 1020

Qy   1198 agtctgagttactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtat 1257
          |     ||| | || ||  | || || |   |   || |  |||||||| |||||
Db   1021 GGGGCAGCCTACGTGGGGGGCATATGCTCCCTGTCCCATGGCGGGGGTGTGAACGAGTAC 1080

Qy   1258 ggtcttccaatggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatc 1317
          ||         |||  ||||        |  | ||||  |||| |  |||||| || ||
Db   1081 GGCAACATGGGGGCGATGGCCGTGACCCTTGCCCAGACGCTGGGACAGAACCTGGGCATG 1140

Qy   1318 caatgggaaccttctagcagaaagccaa---aatgtgactgcacagaatcctggggtggc 1374
             |||  |        | ||   | ||  | ||  | ||   |||  |    ||||   |||
Db   1141 ATGTGGAACAAACACCGGAGCTCGGCAGGGGACTGCAAGTGTCCAGACATCTGGCTGGGC 1200

Qy   1375 tgcatcatggaggaaacaggggtgtcccattctcgaaaattttcaaagtccagcattttg 1434
          ||||||||||||||  || ||| | | ||   | || || || ||   ||||||||
Db   1201 TGCATCATGGAGGACACTGGGTTCTACCTGCCCCGCAAGTTCTCTCGCTGCAGCATCGAC 1260

Qy   1435 gagtatagagacttttttacagagaggaggtggagcctgccttttcaacaggccaacaaag 1494
          |||||  |  | ||| | |||   || ||||||   |||||| ||||||| |||    |||
Db   1261 GAGTACAACCAGTTTCTGCAGGAGGGTGGTGGCAGCTGCCTCTTCAACAAGCCCCTCAAG 1320

Qy   1495 ctatttgagcccacggaatgtggaaatggatacgtggaagctgggggaggagtgtgattgt 1554
```

```
          ||  | || ||| | || || || || || | |||||| || ||||||||||| || ||
Db   1321 CTCCTGGACCCCCCAGAGTGCGGGAACGGCTTCGTGGAGGCAGGGGAGGAGTGCGACTGC 1380

Qy   1555 ggttttcatgtggaatgc---------tatggattatgctgtaagaaatgttccctctcc 1605
          || |       ||| |||           |||    ||||| |||||||  ||||  |
Db   1381 GGCTCGGTGCAGGAGTGCAGCCGCGCAGGTGGCAACTGCTGCAAGAAATGCACCCTGACT 1440

Qy   1606 aacggggctcactgcagcgacgggccctgctgtaacaatacctcatgtctttttcagcca 1665
            |||  ||    |||||||||||||| ||||||      |  |   ||    |  | |||
Db   1441 CACGACGCCATGTGCAGCGACGGGCTCTGCTG------TCGCCGCTGCAAGTACGAACCA 1494

Qy   1666 cgagggtatgaatgccgggatgctgtgaacgagtgtgatattactgaatattgtactgga 1725
          || ||       ||||| || || |||||||||| || ||  | ||    || || ||
Db   1495 CGGGGTGTGTCCTGCCGAGAGGCCGTGAACGAGTGCGACATCGCGGAGACCTGCACCGGG 1554

Qy   1726 gactctggtcagtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaat 1785
          |||||| | |||||||| || || || || |||| |||||  ||    ||  | ||  |
Db   1555 GACTCTAGCCAGTGCCCGCCTAACCTGCACAAGCTGGACGGTTACTACTGTGACCATGAG 1614

Qy   1786 cagggccgctgctacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctgg 1845
          ||||||||||||||    ||    ||||| ||| | |||   ||||| |||     | |||
Db   1615 CAGGGCCGCTGCTACGGAGGTCGCTGCAAAACCCGGGACCGGCAGTGCCAGGTTCTTTGG 1674

Qy   1846 ggaacaaaggctgcagggtctgacaagttctgctatgaaaagctgaatacagaaggcact 1905
          ||       |||||       ||||    |||||||| || |||||||||   || || ||
Db   1675 GGCCATGCGGCTG------CTGATCGCTTCTGCTACGAGAAGCTGAATGTGGAGGGGACG 1728

Qy   1906 gagaagggaaactgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtg 1965
          |||   || | ||| |||     | ||| | | ||| | |||||||| || || || |||
Db   1729 GAGCGTGGGAGCTGTGGGCGCAAGGGATCCGGCTGGGTCCAGTGCAGTAAGCAGGACGTG 1788

Qy   1966 ttctgtggattcttactctgtaccaatcttactcgagctccacgtattggtcaacttcag 2025
          | ||||| ||| | ||||||  |||  |  || |||||||  ||  | ||  | ||  |
Db   1789 CTGTGTGGCTTCCTCCTCTGTGTCAACATCTCTGGAGCTCCTCGGCTAGGGGACCTGGTG 1848

Qy   2026 ggtgagatcattccaacttccttctaccatcaaggccgggtgattgactgcagtggtgcc 2085
          || || ||||. |        |||||||||| || |||  || | | |||||||| || | |
Db   1849 GGAGACATCAGTAGTGTCACCTTCTACCACCAGGGCAAGGAGCTGGACTGCAGGGGAGGC 1908

Qy   2086 catgtagttttagatgatgatacggatgtgggctatgtagaagatggaacgccatgtggc 2145
          || ||      | |  || |    || |||||||  || ||||| ||  | || ||
Db   1909 CACGTGCAGCTGGCGGACGGCTCTGACCTGAGCTATGTGGAGGATGGCACAGCCTGCGGG 1968

Qy   2146 ccgtctatgatgtgtttagatcggaagtgcctacaaattcaagccctaaatatgagcagc 2205
          ||    ||| ||||  | || |      ||||| | |   |   ||| | ||  | |||| |
Db   1969 CCTAACATGTTGTGCCTGGACCATCGCTGCCTGCCAGCTTCTGCCTTCAACTTCAGCACC 2028

Qy   2206 tgtccactcgattccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagcc 2265
          || ||   |  |    || |     | || ||   ||| ||||| || || |||||||
Db   2029 TGCCCCGGCAGTGGGGAGCGCCGGATTTGCTCCCACCACGGGGTCTGCAGCAATGAAGGG 2088

Qy   2266 acctgcatttgtgatttcacctgggcagggacagattgcagtatccgggatccagttagg 2325
          |  ||||| ||| |      |||| |||| | ||| ||||||||||   | ||  |
Db   2089 AAGTGCATCTGTCAGCCAGACTGGACAGGCAAAGACTGCAGTATCCATAACCCCCTGCCC 2148

Qy   2326 aaccttcacccccccaaggat---gaaggacccaagggtcctagtgccaccaatctcata 2382
          |      |  ||| |   |||    ||  ||   || ||||| || | ||||||  ||||
Db   2149 ACGTCCCCACCCACGGGGGAGACGGAGAGATATAAAGGTCCCAGCGGCACCAACATCATC 2208

Qy   2383 ataggctccatcgctggtgccatcctggtagcagctattgtccttgggggcacaggctgg 2442
          || |||||||||||||| ||  ||||||| ||||| || ||||| || ||||||  ||||||
Db   2209 ATTGGCTCCATCGCTGGGGCTGTCCTGGTTGCAGCCATCGTCCTGGGCGGCACGGGCTGG 2268

Qy   2443 ggatttaaaaatgtcaagaagagaaggttcg 2473
          ||||||||||  |        ||||||  ||
Db   2269 GGATTTAAAAACATTCGCCGAGGAAGGTCCG 2299


RESULT  16
E10493
ID   E10493     standard; RNA; HUM; 3183 BP.
XX
AC   E10493;
```

```
XX
SV   E10493.1
XX
DT   08-OCT-1997 (Rel. 52, Created)
DT   02-SEP-2000 (Rel. 65, Last updated, Version 2)
XX
DE   cDNA encoding MDC protein.
XX
KW   JP 1995330799-A/4.
XX
OS   Homo sapiens (human)
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini; Hominidae; Homo.
XX
RN   [1]
RP   1-3183
RA   Nakamura Y., Emi M.;
RT   "MDC PROTEIN AND DNA CODING THE SAME";
RL   Patent number JP1995330799-A/4, 19-DEC-1995.
RL   JAPAN FOUND CANCER RES, EISAI CO LTD.
XX
OS   Homo sapiens (human)
PN   JP 1995330799-A/4
PD   19-DEC-1995
PF   22-APR-1994 JP 1994084470
PR   14-MAY-1993 JP 93P   136602, 22-SEP-1993 JP 93P   257455,
PR   23-FEB-1994 JP 94P    49904, 12-APR-1994 JP 94P    73328
PI   NAKAMURA YUSUKE, EMI MITSURU
PC   C07K16/32,C07K14/82,C12N1/21,C12N15/02,C12N15/09,C12P21/02,
PC   C12P21/08,
PC   C12Q1/68,G01N33/53,G01N33/574,(C12N1/21,C12R1:19),(C12N15/09,
PC   C12R1:91),
PC   (C12P21/02,C12R1:19),(C12P21/08,C12R1:91);
CC   strandedness: Double;
CC   topology: Linear;
CC   hypothetical: No;
CC   anti-sense: No;
FH   Key             Location/Qualifiers
FH
FT   source          1. .3183
FT                   /organism="Homo sapiens"
FT                   /tissue_type="Celebellum"
FT   CDS             1. .2310
FT                   /product="MDC protein"
FT   3'UTR           2311. .3183
XX
FH   Key             Location/Qualifiers
FH
FT   source          1. .3183
FT                   /db_xref="taxon:9606"
FT                   /organism="Homo sapiens"
XX
SQ   Sequence 3183 BP; 645 A; 982 C; 979 G; 577 T; 0 other;


  Query Match             23.0%;  Score 575;  DB 45;  Length 3183;
  Best Local Similarity   57.5%;  Pred. No. 2.3e-127;
  Matches 1226;  Conservative    0;  Mismatches  860;  Indels   45;  Gaps    9;

Qy     376 gaaagcccttatcacgttcttgacacaaaggcaagacaccagcaaaaacataataaggct 435
           ||   ||   | ||    || |||||||| ||   | ||  ||| |  |        | ||
Db     181 GAGGTCCGAAAGCAGCAGCTGGACACAAGGGTCCGCCAGGAGCCACCAGGGGGCCCGCCT 240

Qy     436 gtccatctggcccaggcaagcttccagattgaagccttcggctccaaattcattcttgac 495
           ||||||||||||||||  || |||   ||   |||||||  ||| || ||||  || |||
Db     241 GTCCATCTGGCCCAGGTGAGTTTCGTCATCCCAGCCTTCAACTCAAACTTCACCCTGGAC 300

Qy     496 ctcatactgaacaatggtttgttgtcttctgattatgtggagattcactac------gaa 549
           ||    |||||| |     |  | || ||  | || ||||||   |||| |      ||
Db     301 CTGGAGCTGAACCACCACCTCCTCTCCTCGCAATACGTGGAGCGCCACTTCAGCCGGGAG 360

Qy     550 aatgggaaaccacagtactctaagggtggagagcactgttactaccatggaagcatcaga 609
                 |  |  ||   | |   || |||||| ||||| |||||||| || |   || |

Db     361 GGGACAACCCAGCACAGCACCGGGGCTGGAGACCACTGCTACTACCAGGGGAAGCTCCGG 420

Qy     610 ggcgtcaaagactccaaggtggctctgtcaacctgcaatggacttcatggcatgtttgaa 669
           ||   |  |||||   |  || || || |||||| | || || |||||  | ||
Db     421 GGGAACCCGCACTCCTTCGCCGCCCTCTCCACCTGCCAGGGGCTGCATGGGGTCTTCTCT 480

Qy     670 gatgataccttcgtgtatatgatagagccactagagctggttcatgatgagaaaagcac- 728
           ||||  | |||    || ||  | ||||| | |||| ||| |     |  |  | | |
Db     481 GATGGGAACTTGACTTACATCGTGGAGCCCCAAGAGGTGGCTGGACCTTGGGGAGCCCCT 540

Qy     729 -----aggtcgaccacatataatccagaaaaccttggcaggacagtattctaagcaaatg 783
                |  |  || ||  | ||  |   |||          |  || |      |
Db     541 CAGGGACCCCTTCCCCACCTCATTTACCGGACCCCTCTCCTCCCAGATCCCCTCGGATGC 600

Qy     784 aagaatctcactatggaaagaggtgaccagtggccctttctctctgaattacagtggttg 843
           | | | |                | ||  |   |  |     ||  |  | | |  |
Db     601 AGGGAACCAGGCTGCCTGTTTGCTGTGCCTGCCCAGTCGGCTCCTCCAAACCGGCCGAGG 660

Qy     844 aaaagaaggaagagagcagtgaatccatcacgtggtatatttga---agaaatgaaatat 900
           ||||||||| ||    ||   |    |   || | |  |    ||||  || |||
Db     661 CTGAGAAGGAAAAGGCAGGTCCGCCGGGGCCACCCTACAGTGCACAGTGAAACCAAGTAT 720

Qy     901 ttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgca 960
           ||||  || ||  |  | || || ||| |   ||   || ||   ||    ||    |
Db     721 GTGGAGCTAATTGTGATCAACGACCACCAGCTGTTCGAGCAGATGCGACAGTCGGTGGTC 780

Qy     961 cataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcag 1020
           |  |||| |||||||||  ||||||||||  |||||  |  |||    || ||||||||||
Db     781 CTCACCAGCAACTTTGCCAAGTCCGTGGTGAACCTGGCCGATGTGATATACAAGGAGCAG 840

Qy     1021 ctcaacaccagggttgtcctggtggctgtagagacctggactgagaaggatcagattgac 1080
            |||||||  |  | ||||||||  ||  | || || ||| | ||   |||  ||||  |
Db     841 CTCAACACTCGCATCGTCCTGGTTGCCATGGAAACATGGGCAGATGGGGACAAGATCCAG 900

Qy     1081 atcaccaccaaccctgtgcagatgctccatgagttctcaaaatac---cggcagcgcatt 1137
            |        |||   || |||  ||       | ||       |||   ||| || |  |
Db     901 GTGCAGGATGACCTCCTGGAGACCCTGGCCCGGCTCATGGTCTACCGACGGGAGGGTCTG 960

Qy     1138 aagcagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagc 1197
            |||    | ||||   |||||| ||||| |   ||| || ||    | ||| |||
Db     961 CCTGAGCCCAGTAATGCCACCCACCTCTTCTCGGGCAGGACCTTCCAGAGCACGAGCAGC 1020

Qy     1198 agtctgagttactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtat 1257
            |       ||| | || ||  | || || |   |    || |  |||||||| |||||
Db     1021 GGGGCAGCCTACGTGGGGGGCATATGCTCCCTGTCCCATGGCGGGGGTGTGAACGAGTAC 1080

Qy     1258 ggtcttccaatggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatc 1317
            ||         |||  ||||        |  | ||||  ||||  || |||| || ||
Db     1081 GGCAACATGGGGGCGATGGCCGTGACCCTTGCCCAGACGCTGGGACAGAACCTGGGCATG 1140

Qy     1318 caatgggaaccttctagcagaaagccaa---aatgtgactgcacagaatcctggggtggc 1374
            |||  |        | ||   | ||    | ||  | ||  ||||.  ||||   |||
Db     1141 ATGTGGAACAAACACCGGAGCTCGGCAGGGGACTGCAAGTGTCCAGACATCTGGCTGGGC 1200

Qy     1375 tgcatcatggaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttg 1434
            |||||||||||||| || ||| | | ||    | || || || ||    ||||||||
Db     1201 TGCATCATGGAGGACACTGGGTTCTACCTGCCCCGCAAGTTCTCTCGCTGCAGCATCGAC 1260

Qy     1435 gagtatagagacttttacagagaggaggtggagcctgccttttcaacaggccaacaaag 1494
            ||||| |   | ||| | |||   || |||||   |||||| ||||||| |||    |||
Db     1261 GAGTACAACCAGTTTCTGCAGGAGGGTGGTGGCAGCTGCCTCTTCAACAAGCCCCTCAAG 1320

Qy     1495 ctatttgagcccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgt 1554
            ||  | || ||| | || || || || || | |||||| || ||||||||||| || ||
Db     1321 CTCCTGGACCCCCCAGAGTGCGGGAACGGCTTCGTGGAGGCAGGGGAGGAGTGCGACTGC 1380

Qy     1555 ggttttcatgtggaatgc---------tatggattatgctgtaagaaatgttccctctcc 1605
            || |       ||| |||           |||    ||||| ||||||||  |||| |
Db     1381 GGCTCGGTGCAGGAGTGCAGCCGCGCAGGTGGCAACTGCTGCAAGAAATGCACCCTGACT 1440

Qy     1606 aacggggctcactgcagcgacgggccctgctgtaacaatacctcatgtctttttcagcca 1665
            |||  ||    |||||||||||||| ||||||      |   |   ||    |  | |||
Db     1441 CACGACGCCATGTGCAGCGACGGGCTCTGCTG------TCGCCGCTGCAAGTACGAACCA 1494
```

```
Qy   1666 cgagggtatgaatgccgggatgctgtgaacgagtgtgatattactgaatattgtactgga 1725
          || ||       |||||  || || ||||||||||| || ||  | ||    || || ||
Db   1495 CGGGGTGTGTCCTGCCGAGAGGCCGTGAACGAGTGCGACATCGCGGAGACCTGCACCGGG 1554

Qy   1726 gactctggtcagtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaat 1785
          |||||| | |||||||| || || || ||  |||| ||||| ||    ||  | ||  |
Db   1555 GACTCTAGCCAGTGCCCGCCTAACCTGCACAAGCTGGACGGTTACTACTGTGACCATGAG 1614

Qy   1786 cagggccgctgctacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctgg 1845
          ||||||||||||||   ||    ||||| ||| | |||   ||||| |||    | |||
Db   1615 CAGGGCCGCTGCTACGGAGGTCGCTGCAAAACCCGGGACCGGCAGTGCCAGGTTCTTTGG 1674

Qy   1846 ggaacaaaggctgcagggtctgacaagttctgctatgaaaagctgaatacagaaggcact 1905
          ||      |||||      ||||    |||||||| || |||||||||   || || ||
Db   1675 GGCCATGCGGCTG------CTGATCGCTTCTGCTACGAGAAGCTGAATGTGGAGGGGACG 1728

Qy   1906 gagaagggaaactgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtg 1965
          |||   || | ||| |||    | |||  | | ||| | |||||||| || || || |||
Db   1729 GAGCGTGGGAGCTGTGGGCGCAAGGGATCCGGCTGGGTCCAGTGCAGTAAGCAGGACGTG 1788

Qy   1966 ttctgtggattcttactctgtaccaatcttactcgagctccacgtattggtcaacttcag 2025
          | ||||| ||| | |||||  |||  |  || |||||||| ||  | ||  | ||   |
Db   1789 CTGTGTGGCTTCCTCCTCTGTGTCAACATCTCTGGAGCTCCTCGGCTAGGGGACCTGGTG 1848

Qy   2026 ggtgagatcattccaacttccttctaccatcaaggccgggtgattgactgcagtggtgcc 2085
          || || |||| |       ||||||||||| || |||  || | | |||||||| || | |
Db   1849 GGAGACATCAGTAGTGTCACCTTCTACCACCAGGGCAAGGAGCTGGACTGCAGGGGAGGC 1908

Qy   2086 catgtagttttagatgatgatacggatgtgggctatgtagaagatggaacgccatgtggc 2145
          || ||     | |  || |   | ||  || ||||||| || |||||  ||   | || ||
Db   1909 CACGTGCAGCTGGCGGACGGCTCTGACCTGAGCTATGTGGAGGATGGCACAGCCTGCGGG 1968

Qy   2146 ccgtctatgatgtgtttagatcggaagtgcctacaaattcaagccctaaatatgagcagc 2205
          ||    ||| ||||  | || |     |||||  | |  |   ||| | ||  | ||||  |
Db   1969 CCTAACATGTTGTGCCTGGACCATCGCTGCCTGCCAGCTTCTGCCTTCAACTTCAGCACC 2028

Qy   2206 tgtccactcgattccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagcc 2265
          || ||   |  |    || |     | || ||   ||| ||||| || || |||||||
Db   2029 TGCCCCGGCAGTGGGGAGCGCCGGATTTGCTCCCACCACGGGGTCTGCAGCAATGAAGGG 2088

Qy   2266 acctgcatttgtgatttcacctgggcagggacagattgcagtatccggatccagttagg 2325
          |  ||||| ||| |      |||| |||| | ||| ||||||||||   | ||  |
Db   2089 AAGTGCATCTGTCAGCCAGACTGGACAGGCAAAGACTGCAGTATCCATAACCCCCTGCCC 2148

Qy   2326 aaccttcacccccccaaggat---gaaggacccaagggtcctagtgccaccaatctcata 2382
          |     |  ||| |   |||    ||  ||   || ||||| || | |||,||  ||||
Db   2149 ACGTCCCCACCCACGGGGGAGACGGAGAGATATAAAGGTCCCAGCGGCACCAACATCATC 2208

Qy   2383 ataggctccatcgctggtgccatcctggtagcagctattgtccttgggggcacaggctgg 2442
          || |||||||||||||| ||  ||||||| ||||| || ||||| || ||||| ||||||
Db   2209 ATTGGCTCCATCGCTGGGGCTGTCCTGGTTGCAGCCATCGTCCTGGGCGGCACGGGCTGG 2268

Qy   2443 ggatttaaaaatgtcaagaagagaaggttcg 2473
          ||||||||||  |       |||||| ||
Db   2269 GGATTTAAAAACATTCGCCGAGGAAGGTCCG 2299
```

```
RESULT  17
AB009676
LOCUS       AB009676     3229 bp    mRNA            ROD       17-AUG-1999
DEFINITION  Mus musculus mRNA for ADAM11, complete cds.
ACCESSION   AB009676
VERSION     AB009676.1  GI:5736633
KEYWORDS    ADAM11.
SOURCE      Mus musculus brain cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Sagane,K., Yamazaki,K., Mizui,Y. and Tanaka,I.
  TITLE     Cloning and chromosomal mapping of mouse ADAM11, ADAM22 and ADAM23
  JOURNAL   Gene 236 (1), 79-86 (1999)
  MEDLINE   99365303
REFERENCE   2  (bases 1 to 3229)
  AUTHORS   Sagane,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-DEC-1997) to the DDBJ/EMBL/GenBank databases. Koji
            Sagane, Eisai Co.,Ltd., Tsukuba Research Laboratories; Tokodai
            5-1-3, Tsukuba, Ibaraki 300-2635, Japan
            (E-mail:k-sagane@hhc.eisai.co.jp, Tel:+81-298-47-5813,
            Fax:+81-298-47-5367)
FEATURES             Location/Qualifiers
     source          1. .3229
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /tissue_type="brain"
     CDS             16. .2337
                     /codon_start=1
                     /product="ADAM11"
                     /protein_id="BAA83384.1"
                     /db_xref="GI:5736634"
                     /translation="MRRLRRWAIAALLLLPLLPPPGLGALGPRGALHWRSSAHVGSPE
                     SPEGSEVTEPSRLVRQSSGGEVRKPQLDTRVRQDPPRGTPVHLAQVSFVIPAFDSNFT
                     LDLELNHHLLSSQYVERHFSREGTRQHSTGAGDHCYYHGKLRGNPQSFAALSTCQGLH
                     GVFSDGNLTYIVEPKEIAGPRGPPQGPLPHLIYRTPLLPAPLGCREPGCLFAVPAQST
                     LPNWPKLRRKRQVRRGHPTVHSETKYVELIVINDHQLFEQMRQSVVLTSNFAKSVVNL
                     ADVIYKEQLNTRIVLVAMETWADGDKIQVQDDLLETLARLMVYRREGLPEPSDATHLF
                     SGRTFKSTSSGAAYVGGICSLSRGGGVNEYGNMGAMAVTLAQTLGQNLGMMWNKHRSS
                     AGDCKCPDIWLGCIMEDTGFYLPRKFSRCSIDEYNQFLQEGGGSCLFNKPLKLLDPPE
                     CGNGFVEAGEECDCGSVQECSRAGGNCCKKCTLTHDAMCSDGLCCRRCKYEPRGVSCR
                     EAVNECDIAETCTGDSSQCPPNLHKLDGYYCDHEQGRCYGGRCKTRDRQCQALWGHAA
                     ADRFCYEKLNVEGTERGNCGRKGSGWVQCSKQDVLCGFLLCVNISGAPRLGDLGGDIS
                     SVTFYHQGKELDCRGGHVQLADGSDLSYVEDGTACGPNMLCLDHRCLPASAFNFSTCP
                     GSGERRICSHHGVCSNEGKCICQPDWTGKDCSIHNPLPTSPPTGETERYKGPSGTNII
                     IGSIAGAVLVAAIVLGGTGWGFKNIRRGRSGGA"
     polyA_site      3229
                     /note="20 a nucleotides"
BASE COUNT      650 a    966 c    952 g    649 t     12 others
ORIGIN
```

```
  Query Match             22.5%;  Score 562.8;  DB 94;  Length 3229;
  Best Local Similarity   57.1%;  Pred. No. 2e-124;
  Matches 1205;  Conservative    0;  Mismatches  862;  Indels   45;  Gaps    8;

Qy    395 ttgacacaaaggcaagacaccagcaaaaacataataaggctgtccatctggcccaggcaa 454
          | ||||| | ||   | ||  | |           | | |||| || ||||||||||  |
Db    227 TGGACACCAGGGTCCGCCAGGATCCGCCCAGGGGGACGCCTGTGCACCTGGCCCAGGTGA 286

Qy    455 gcttccagattgaagccttcggctccaaattcattcttgacctcatactgaacaatggtt 514
          | |||   ||    ||||||| ||| || |||| ||| |||||    |||||| ||
Db    287 GTTTCGTCATCCCGGCCTTCGACTCAAACTTCACTCTGGACCTGGAGCTGAACCATCACC 346

Qy    515 tgttgtcttctgattatgtggagattcactac------gaaaatgggaaaccacagtact 568
          |  |||| ||  | ||||||||||   |||| |       ||       | || |||   |
Db    347 TCCTGTCCTCGCAGTATGTGGAGCGCCACTTCAGCCGGGAGGGAACAAGACAACACAGCA 406

Qy    569 ctaagggtggagagcactgttactaccatggaagcatcagaggcgtcaaagactccaagg 628
          ||  || ||||||  ||||| ||||||||||| |   || | |||  |  | | |||   |
Db    407 CTGGGGCTGGAGACCACTGCTACTACCATGGGAAACTCCGGGGCAACCCACAGTCCTTTG 466

Qy    629 tggctctgtcaacctgcaatggacttcatggcatgtttgaagatgataccttcgtgtata 688
            || || || || ||| | || || |||||  | ||    ||||  | | |    || |
Db    467 CTGCACTCTCTACATGCCAGGGGCTGCATGGGGTCTTCTCTGATGGCAACCTGACTTACA 526

Qy    689 tgatagagccactagagctggttcatgatgagaaaagcacaggtcgaccacat------- 741
          |  |||||||    ||| | | |         |  |  | ||    |||| | |
Db    527 TCGTAGAGCCTAAGGAGATAGCTGGGCCCCGGGGACCCCCACAGGGACCCCTTCCCCACC 586

Qy    742 --ataatccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcactatgg 799
            ||   || ||   |  |       |       |       | | ||     |    ||
Db    587 TCATTTACCGGACCCCTCTCCTCCCAGCCCCCCTTGGATGCAGGGAGCCAGGCTGCCTGT 646

Qy    800 aaagaggtgaccagtggccctttctctctgaattacagtggttgaaaagaaggaagagag 859
                |          | |   |||| |  | |  |    | | | ||| |  | |  |
```

```
Db    647 TTGCTGTCCCTGCCCAGTCTACTCTCCCCAACTGGCCCAAGCTAAGAAGGAAAAGGCAGG 706

Qy    860 cagtgaatccatcacgtggtatatttgaagaaatgaaatatttggaacttatgattgtta 919
Db    707 TCCGCAGGGGCCACCCCACAGTGCACAGCGAGACCAAGTATGTGGAGTTGATTGTAATCA 766

Qy    920 atgatcacaaaacgtataagaagcatcgctcttctcatgcacatacccaacaactttgcaa 979
Db    767 ATGACCACCAGCTGTTTGAGCAGATGCGGCAGTCAGTGGTCCTCACCAGCAACTTTGCAA 826

Qy    980 agtccgtggtcaaccttgtggattctatttacaaggagcagctcaacaccagggttgtcc 1039
Db    827 AATCTGTTGTGAACCTGGCAGATGTGATATACAAGGAACAGCTCAACACAAGAATTGTTC 886

Qy   1040 tggtggctgtagagacctggactgagaaggatcagattgacatcaccaccaaccctgtgc 1099
Db    887 TGGTTGCCATGGAAACGTGGGCAGATGGGGACAAGATCCAGGTGCAGGATGACCTACTGG 946

Qy   1100 agatgctccatgagttctcaaaataccggc---agcgcattaagcagcatgctgatgctg 1156
Db    947 AGACCCTGGCCCGGCTTATGGTCTACCGGCGGGAGGGTCTGCCTGAGCCCAGTGATGCCA 1006

Qy   1157 tgcacctcatctcgcgggtgacatttcactataagagaagcagtctgagttactttggag 1216
Db   1007 CCCACCTCTTCTCGGGTAGAACCTTCAAGAGCACCAGCAGCGGGGCGGCCTACGTGGGAG 1066

Qy   1217 gtgtctgttctcgcacaagaggagttggtgtgaatgagtatggtcttccaatggcagtgg 1276
Db   1067 GCATCTGTTCACTGTCCCGGGGTGGAGGTGTGAACGAGTATGGCAACATGGGTGCCATGG 1126

Qy   1277 cacaagtattatcgcagagcctggctcaaaaccttggaatccaatgggaaccttctagca 1336
Db   1127 CGGTGACCCTGGCCCAGACGCTAGGGCAGAACTTGGGCATGATGTGGAATAAGCACCGGA 1186

Qy   1337 gaaagccaa---aatgtgactgcacagaatcctggggtggctgcatcatggaggaaacag 1393
Db   1187 GCTCAGCAGGGGACTGCAAGTGTCCAGACATTTGGCTGGGCTGCATCATGGAGGACACTG 1246

Qy   1394 gggtgtcccattctcgaaaattttcaaagtgcagcattttggagtatagagacttttac 1453
Db   1247 GGTTCTATTTGCCCCGCAAGTTCTCGCGCTGCAGCATCGACGAATACAACCAGTTTCTGC 1306

Qy   1454 agagaggaggtggagcctgccttttcaacaggccaacaaagctatttgagcccacggaat 1513
Db   1307 AGGAGGGAGGCGGGAGCTGCCTGTTCAACAAGCCCCTCAAGCTTCTGGACCCTCCCGAGT 1366

Qy   1514 gtggaaatggatacgtggaagctggggaggagtgtgattgtggttttcatgtggaatgct 1573
Db   1367 GCGGAAACGGCTTCGTGGAGGCGGGAGAGGAGTGCGACTGCGGGTCGGTGCAGGAGTGCA 1426

Qy   1574 atggatta---------tgctgtaagaaatgttccctctccaacggggctcactgcagcg 1624
      427 GCCGAGCAGGGGGCAACTGCTGCAAAAAATGCACCTTGACGCACGACGCCATGTGCAGCG 1486

      625 acgggccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgccggg 1684
Db   1487 ATGGGCTCTGTTG------TCGCCGCTGCAAGTATGAGCCACGAGGTGTCTCCTGCCGAG 1540

Qy   1685 atgctgtgaacgagtgtgatattactgaatattgtactggagactctggtcagtgcccac 1744
Db   1541 AAGCGGTGAATGAGTGTGACATTGCAGAGACCTGCACCGGCGACTCAAGCCAGTGTCCCC 1600

Qy   1745 caaatcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctacaatg 1804
Db   1601 CTAACCTTCACAAGCTGGACGGTTACTACTGTGATCATGAGCAGGGTCGTTGCTATGGAG 1660

Qy   1805 gcgagtgcaagaccagagacaaccagtgtcagtacatctggggaacaaaggctgcagggt 1864
Db   1661 GCCGCTGTAAAACCCGGGACCGGCAGTGCCAAGCCCTATGGGGCCATGCGGCTG------ 1714

Qy   1865 ctgacaagttctgctatgaaaagctgaatacagaaggcactgagaagggaaactgcggga 1924
Db   1715 CGGATCGTTTCTGCTATGAGAAGCTGAACGTGGAGGGGACAGAGCGTGGAAACTGTGGAC 1774

Qy   1925 aggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtggattcttactct 1984
Db   1775 GCAAGGGATCTGGTTGGGTCCAGTGCAGTAAGCAGGATGTGCTCTGTGGCTTCCTTCTGT 1834

Qy   1985 gtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatcattccaactt 2044
Db   1835 GCGTCAACATCTCTGGAGCTCCGCGGCTAGGGGATCTGGGGGGCGACATCAGCAGTGTCA 1894

Qy   2045 ccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagttttagatgatg 2104
Db   1895 CCTTCTACCACCAGGGCAAGGAGTTGGACTGCAGGGGAGGCCACGTGCAGCTAGCTGATG 1954

Qy   2105 atacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatgatgtgtttag 2164
Db   1955 GCTCGGACCTGAGCTATGTGGAGGACGGCACGGCCTGTGGGCCCAACATGTTGTGCCTAG 2014

Qy   2165 atcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactcgattccaagg 2224
Db   2015 ATCACCGCTGCCTGCCAGCCTCTGCCTTCAACTTCAGCACCTGCCCTGGAAGTGGAGAGC 2074

Qy   2225 gtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatttgtgatttca 2284
Db   2075 GAAGGATCTGCTCCCATCATGGGGTTTGCAGCAACGAGGGGAAGTGTATCTGTCAGCCAG 2134

Qy   2285 cctgggcagggacagattgcagtatccgggatccagtt---aggaaccttcaccccccca 2341
Db   2135 ACTGGACAGGCAAAGACTGCAGTATTCACAACCCACTGCCCACGTCCCCTCCCACTGGGG 2194

Qy   2342 aggatgaaggacccaagggtcctagtgccaccaatctcataataggctccatcgctggtg 2401
Db   2195 AGACTGAGAGATACAAAGGTCCCAGCGGTACCAACATCATCATTGGCTCCATCGCCGGGG 2254

Qy   2402 ccatcctggtagcagctattgtccttgggggcacaggctggggatttaaaaatgtcaaga 2461
Db   2255 CTGTCCTGGTCGCAGCCATCGTCCTGGGCGGCACGGGCTGGGGATTTAAAAACATCCGTC 2314

Qy   2462 agagaaggttcg 2473
Db   2315 GTGGAAGGTCCG 2326


RESULT  18
AF032383
LOCUS       AF032383     3447 bp    mRNA            VRT       30-SEP-1998
DEFINITION  Xenopus laevis metalloprotease-disintegrin (MDC11b) mRNA, complete
            cds.
ACCESSION   AF032383
VERSION     AF032383.1  GI:3510508
KEYWORDS    .
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 3447)
  AUTHORS   Cai,H., Kraetzschmar,J., Alfandari,D., Hunnicutt,G. and Blobel,C.P.
  TITLE     Neural crest-specific and general expression of distinct
            metalloprotease-disintegrins in early Xenopus laevis development
  JOURNAL   Dev. Biol. (1998) In press
REFERENCE   2  (bases 1 to 3447)
  AUTHORS   Cai,H., Kraetzschmar,J. and Blobel,C.B.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-1997) CB+G, Cornell University Medical College,
            New York, NY 10021, USA
FEATURES             Location/Qualifiers
     source          1. .3447
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
     gene            1. .3447
                     /gene="MDC11b"
     CDS             139. .2946
                     /gene="MDC11b"
```

```
                     /codon_start=1
                     /product="metalloprotease-disintegrin"
                     /protein_id="AAC61847.1"
                     /db_xref="GI:3510509"
                     /translation="MHINGGPLASWICCVIGSIHLAHASTRPENGGTSGMQRKKENSV
                     LGMEDTVPLRLIFSNEEDNQTTQGLLSTRVRAGSPQHQDQLTHVAQASFQIDAFGSSF
                     ILDVELNHDLLSSDYRERHVTQDGKTVEVKGGEHCYYQGQIRGKAKSFVALSTCNGLH
                     GMFCDGNHTYLIEPGEKYNPNEDYQFHSVYKSKVLEFPLDELPSEFWALNDTSVRLSQ
                     QTRSQRKKRQTRRYPRNVEDETKYVELMIVNDHLMYKKHRLSVGHTNSYAKSVVNMAD
                     LIYKEQLNTRIVLVAMETWATDNKFSISENPLLTLKEFMKYRRDFIKDKSDAVHLFSG
                     SQLRVAAVVLRILVEWCSLLKGGGVNEFGKPDVMAVTLAQSLAHNLGIFSDKKKLLSG
                     ECKCEDTWSGCIMGDIGYYLPSKFSVCNIEEYHEFLNNGGGACLFNKPLKLLDPPECG
                     NGFVETGEECDCGTIAECAMEGEECCKKCTLTQDSECSDGLCCSNCKFNPKEMLCREA
                     VNDCDIPETCTGNTSQCPANIHKLDGYSCESMQGLCFGGRCKTRDRQCKYIWGEKVSA
                     ADRYCYEKLNIEGTEKGNCGRNKETWIQCNKQDVLCGYLLCTNISNVPRLGELDGDVT
                     SSSIVNQGKLYNCSGGHVKLDEDTDLGYVEDGTPCGTGMMCLEHRCLPIDSFNFSTCL
                     GSTNKICSGHGVCSNEVRCICDRFWTGEDCSSYLHYDHIKPEGDNRDEGVISTNIIIG
                     AIAGTILVLALVLGITAWGYKNYRRERQIPQGDYVKKPGDADSFYSDLPPGVSSNSAS
                     SSKKRSAILSHFQISACSIPHYSISQNISLFCRRSNGLSHSWSERIPDTKHVSDVCEN
                     GRPRSNSWQGNVTSSRKKLRGKRFRPRSNSTETLSPAKSPSSSTGSIASSRRYPYPMP
                     PLPDEERKASKQSARLWETSI"
BASE COUNT     1050 a    729 c    834 g    834 t
ORIGIN


  Query Match           21.6%;  Score 540.6;  DB 8;  Length 3447;
  Best Local Similarity 56.3%;  Pred. No. 4.1e-119;
  Matches 1165;  Conservative    0;  Mismatches  874;  Indels   30;  Gaps   7;

Qy    406 gcaagacaccagcaaaaacataataaggctgtccatctggcccaggcaagcttccagatt 465
Db    364 GCAGGGAGCCCTCAGCACCAGGATCAGCTCACTCACGTTGCCCAGGCCAGCTTTCAGATT 423

Qy    466 gaagccttcggctccaaattcattcttgacctcatactgaacaatggcttgttgtcttct 525
Db    424 GATGCATTTGGGTCTTCCTTCATCCTGGATGTGGAACTCAATCATGACTTGTTGTCTTCA 483

Qy    526 gattatgtggagattca---ctacgaaaatgggaaaccacagtactcaaagggtggagag 582
Db    484 GACTACCGGGAAAGACATGTAACCCAGGATGGGAAAACAGTGGAGGTCAAAGGCGGAGAA 543

Qy    583 cactgttactaccatggaagcatcagaggcgtcaaagactccaaggtggctctgtcaacc 642
Db    544 CATTGTTATTACCAAGGCCAAATAAGAGGAAAAGCCAAATCTTTTGTGGCTCTATCAACT 603

      643 tgcaatggacttcatggcatgtttgaagatgataccttcgtgtatatgatagagccacta 702
Db    604 TGCAATGGACTTCATGGGATGTTCTGTGATGGAAACCACACATATCTTATTGAACCAGGA 663

Qy    703 gagctggttcatgatgagaaaagcacaggtcgaccacatataatccacaaaaccttggca 762
Db    664 GAGAAGTACAACCCAAATGAGGACTATCAGTTCCATTCTGTATATAAATCAAAAGTACTT 723

Qy    763 ggacagtattctaagcaaatgaagaatctcactatggaaagaggtgaccagtggcccttt 822
Db    724 GAATTTCCTCTGGATGAGCTGCCATCTGAATTCTGGGCACTGAATGACACTTCAGTAAGG 783

Qy    823 ctctctgaattacagtggttgaaaagaaggaagagag---cagtgaatccatcacgtggt 879
Db    784 CTAAGCCAACAGACTAGAAGCCAAAGGAAAAAAAGGCAGACACGGCGGTATCCTCGCAAT 843

Qy    880 atatttgaagaaatgaaatatttggaacttatgattgttaatgatcacaaaacgtataag 939
Db    844 GTCGAAGATGAGACCAAATATGTTGAGCTTATGATCGTTAATGACCATTTAATGTATAAA 903

Qy    940 aagcatcgctcttctcatgcacataccaacaactttgcaaagtccgtggtcaaccttgtg 999
Db    904 AAGCACCGTTTATCAGTTGGTCACACAAACAGCTATGCCAAATCCGTTGTGAACATGGCA 963

Qy   1000 gattctatttacaaggagcagctcaacaccagggttgtcctggtggctgtagagacctgg 1059
Db    964 GATTTAATATACAAGGAGCAGTTAAACACAAGGATAGTGCTGGTGGCCATGGAAACCTGG 1023

Qy   1060 actgagaaggatcagattgacatcaccaccaaccctgtgcagatgctccatgagttctca 1119
Db   1024 GCAACTGATAACAAGTTCAGCATCTCTGAAAATCCATTGCTAACTTTGAAGGAATTCATG 1083

Qy   1120 aaatac---cggcagcgcattaagcagcatgctgatgctgtgcacctcatctcgcgggtg 1176
Db   1084 AAATACAGACGGGATTTCATCAAAGATAAAAGTGATGCTGTCCACCTTTTTTCGGGAAGT 1143

Qy   1177 acatttcactataagagaagcagtctgagttactttggaggtgtctgttctcgcacaaga 1236
Db   1144 CAGTTGAGAGTAGCCGCAGTGGTGCTGCGTATATTGGTGGAGTGGTGCTCACTGTTAAAA 1203

Qy   1237 ggagttggtgtgaatgagtatggtcttccaatggcagtggcacaagtattatcgcagagc 1296
Db   1204 GGAGGAGGTGTGAATGAGTTTGGAAAACCGGATGTAATGGCCGTTACTTTGGCTCAGTCT 1263

Qy   1297 ctggctcaaaaccttggaatccaatgggaaccttcta---gcagaaagccaaaatgtgac 1353
Db   1264 CTGGCTCATAACCTCGGAATATTCTCAGACAAGAAGAAGTTGTTAAGTGGTGAATGCAAG 1323

Qy   1354 tgcacagaatcctggggtggctgcatcatggaggaaacaggggtgtcccattctcgaaaa 1413
Db   1324 TGTGAGGATACGTGGTCTGGATGCATCATGGGAGATATTGGGTATTACCTTCCAAGCAAG 1383

Qy   1414 ttttcaaagtgcagcattttggagtatagagacttttttacagagaggaggtggagcctgc 1473
Db   1384 TTCTCCGTTTGTAACATAGAAGAATACCATGAATTTCTAAATAACGGGGGAGGAGCTTGC 1443

Qy   1474 cttttcaacaggccaacaaagctatttgagcccacggaatgtggaaatggatacgtggaa 1533
Db   1444 CTGTTTAATAAGCCACTTAAGCTTTTGGATCCTCCCGAGTGTGGAAATGGCTTTGTTGAG 1503

Qy   1534 gctggggaggagtgtgattgtggttttcatgtggaatgctat---------ggattatgc 1584
Db   1504 ACCGGGGAAGAGTGTGATTGTGGCACTATTGCAGAATGTGCTATGGAAGGAGAGGAGTGC 1563

Qy   1585 tgtaagaaatgttccctctccaacggggctcactgcagcgacgggccctgctgtaacaat 1644
Db   1564 TGCAAAAAGTGCACCTTAACTCAAGACTCAGAGTGCAGTGATGGGCTGTGCTGTTCTAA- 1622

Qy   1645 acctcatgtctttttcagccacgagggtatgaatgccgggatgctgtgaacgagtgtgat 1704
Db   1623 -----CTGTAAGTTCAATCCTAAAGAGATGCTGTGCCGAGAAGCTGTAAACGATTGTGAT 1677

Qy   1705 attactgaatattgtactggagactctggtcagtgcccaccaaatcttcataagcaagac 1764
Db   1678 ATCCCAGAGACCTGTACAGGAAATACAAGCCAGTGTCCTGCTAACATTCACAAACTGGAT 1737

Qy   1765 ggatatgcatgcaatcaaaatcagggccgctgctacaatggcgagtgcaagaccagagac 1824
Db   1738 GGATACTCTTGTGAAAGCATGCAGGGTCTCTGTTTTGGAGGAAGATGCAAAACAAGAGAC 1797

Qy   1825 aaccagtgtcagtacatctggggaacaaaggctgcagggtctgacaagttctgctatgaa 1884
Db   1798 CGACAGTGCAAGTATATCTGGGGTGAAAAAGTGTCGGCTGCAGACAGATATTGCTATGAA 1857

Qy   1885 aagctgaatacagaaggcactgagaagggaaactgcgggaaggatggagaccggtggatt 1944
Db   1858 AAGCTGAACATTGAAGGCACAGAAAAAGGCAACTGTGGTAGAAATAAGGAAACCTGGATA 1917

Qy   1945 cagtgcagcaaacatgatgtgttctgtggattcttactctgtaccaatcttactcgagct 2004
Db   1918 CAGTGCAACAAACAAGATGTACTATGTGGCTATCTCTTGTGCACCAATATTTCCAATGTG 1977

Qy   2005 ccacgtattggtcaacttcagggtgagatcattccaacttccttctaccatcaaggccgg 2064
Db   1978 CCGAGGCTGGGGGAGCTGGATGGAGATGTTACCTCTTCTTCAATTGTAAATCAGGGAAAG 2037

Qy   2065 gtgattgactgcagtggtgcccatgtagttttagatgatgatacggatgtgggctatgta 2124
Db   2038 CTGTATAACTGCAGTGGTGGGCATGTGAAGCTGGATGAAGACACCGACCTTGGCTATGTA 2097

Qy   2125 gaagatggaacgccatgtggcccgtctatgatgtgtttagatcggaagtgcctacaaatt 2184
```

```
Db   2098 GAAGATGGCACTCCTTGTGGCACTGGCATGATGTGCCTTGAACACAGATGCCTTCCCATC 2157

Qy   2185 caagccctaaatatgagcagctgtccactcgattccaagggtaaagtctgttcgggccat 2244
Db   2158 GATTCCTTCAACTTCAGCACCTGCTTGGGCAGCACAAA---CAAAATTTGCTCAGGGCAT 2214

Qy   2245 ggggtgtgtagtaatgaagccacctgcatttgtgatttcacctgggcagggacagattgc 2304
Db   2215 GGTGTATGTAGCAACGAGGTTAGGTGCATCTGCGACAGGTTTTGGACAGGAGAAGACTGC 2274

Qy   2305 agtatccgggatccagttaggaaccttcacccccccaaggatgaaggacccaagggtcct 2364
Db   2275 AGCAGTTACTTACATTATGATCATATTAAACCGGAAGGAGACAACAGAGATGAAGGTGTA 2334

Qy   2365 agtgccaccaatctcataataggctccatcgctggtgccatcctggtagcagctattgtc 2424
Db   2335 ATTAGCACAAACATAATAATTGGAGCTATTGCTGGCACCATACTAGTGCTGGCCCTGGTG 2394

Qy   2425 cttgggggcacaggctggggatttaaaaa 2453
Db   2395 TTAGGAATAACCGCATGGGGGTACAAAAA 2423


RESULT  19
I25847
LOCUS       I25847     2913 bp    DNA             PAT       07-OCT-1996
DEFINITION  Sequence 7 from patent US 5552526.
ACCESSION   I25847
VERSION     I25847.1  GI:1605717
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 2913)
  AUTHORS   Nakamura,Y. and Emi,M.
  TITLE     MDC proteins and DNAS encoding the same
  JOURNAL   Patent: US 5552526-A 7 03-SEP-1996;
FEATURES             Location/Qualifiers
     source          1. .2913
                     /organism="unknown"
BASE COUNT      604 a    888 c    885 g    536 t
ORIGIN


  Query Match             21.2%;  Score 530;  DB 10;  Length 2913;
  Best Local Similarity   57.8%;  Pred. No. 1.5e-116;
  Matches 1114;  Conservative    0;  Mismatches  775;  Indels   39;  Gaps    8;

Qy    573 gggtggagagcactgttactaccatggaagcatcagaggcgtcaaagactccaaggtggc 632
Db    114 GGCTGGAGACCACTGCTACTACCAGGGGAAGCTCCGGGGGAACCCGCACTCCTTCGCCGC 173

Qy    633 tctgtcaacctgcaatggacttcatggcatgtttgaagatgataccttcgtgtatatgat 692
Db    174 CCTCTCCACCTGCCAGGGGCTGCATGGGGTCTTCTCTGATGGGAACTTGACTTACATCGT 233

Qy    693 agagccactagagctggttcatgatgagaaaagcac------aggtcgaccacatataat 746
Db    234 GGAGCCCCAAGAGGTGGCTGGACCTTGGGGAGCCCCTCAGGGACCCCTTCCCCACCTCAT 293

Qy    747 ccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcactatggaaagagg 806
Db    294 TTACCGGACCCCTCTCCTCCCAGATCCCCTCGGATGCAGGGAACCAGGCTGCCTGTTTGC 353

Qy    807 tgaccagtggccctttctctctgaattacagtggttgaaaagaaggaagagagcagtgaa 866
Db    354 TGTGCCTGCCCAGTCGGCTCCTCCAAACCGGCCGAGGCTGAGAAGGAAAAGGCAGGTCCG 413

Qy    867 tccatcacgtggtatatttga---agaaatgaaatatttggaacttatgattgttaatga 923
Db    414 CCGGGGCCACCCTACAGTGCACAGTGAAACCAAGTATGTGGAGCTAATTGTGATCAACGA 473

Qy    924 tcacaaaacgtataagaagcatcgctcttctcatgcacataccaacaactttgcaaagtc 983
Db    474 CCACCAGCTGTTCGAGCAGATGCGACAGTCGGTGGTCCTCACCAGCAACTTTGCCAAGTC 533

Qy    984 cgtggtcaaccttgtggattctatttacaaggagcagctcaacaccagggttgtcctggt 1043
Db    534 CGTGGTGAACCTGGCCGATGTGATATACAAGGAGCAGCTCAACACTCGCATCGTCCTGGT 593

Qy   1044 ggctgtagagacctggactgagaaggatcagattgacatcaccaccaaccctgtgcagat 1103
Db    594 TGCCATGGAAACATGGGCAGATGGGGACAAGATCCAGGTGCAGGATGACCTCCTGGAGAC 653

Qy   1104 gctccatgagttctcaaaatac---cggcagcgcattaagcagcatgctgatgctgtgca 1160
Db    654 CCTGGCCCGGCTCATGGTCTACCGACGGGAGGGTCTGCCTGAGCCCAGTAATGCCACCCA 713

Qy   1161 cctcatctcgcgggtgacatttcactataagagaagcagtctgagttactttggaggtgt 1220
Db    714 CCTCTTCTCGGGCAGGACCTTCCAGAGCACGAGCAGCGGGGCAGCCTACGTGGGGGGCAT 773

Qy   1221 ctgttctcgcacaagaggagttggtgtgaatgagtatggtcttccaatggcagtggcaca 1280
Db    774 ATGCTCCCTGTCCCATGGCGGGGGTGTGAACGAGTACGGCAACATGGGGGCGATGGCCGT 833

Qy   1281 agtattatcgcagagcctggctcaaaaccttggaatccaatgggaaccttctagcagaaa 1340
Db    834 GACCCTTGCCCAGACGCTGGGACAGAACCTGGGCATGATGTGGAACAAACACCGGAGCTC 893

Qy   1341 gcca---aaatgtgactgcacagaatcctggggtggctgcatcatggaggaaacaggggt 1397
Db    894 GGCAGGGGACTGCAAGTGTCCAGACATCTGGCTGGGCTGCATCATGGAGGACACTGGGTT 953

Qy   1398 gtcccattctcgaaaattttcaaagtgcagcattttggagtatagagacttttttacagag 1457
Db    954 CTACCTGCCCCGCAAGTTCTCTCGCTGCAGCATCGACGAGTACAACCAGTTTCTGCAGGA 1013

Qy   1458 aggaggtggagcctgccttttcaacaggccaacaaagctatttgagcccacggaatgtgg 1517
Db   1014 GGGTGGTGGCAGCTGCCTCTTCAACAAGCCCCTCAAGCTCCTGGACCCCCCAGAGTGCGG 1073

Qy   1518 aaatggatacgtggaagctggggaggagtgtgattgtggttttcatgtggaatgc----- 1572
Db   1074 GAACGGCTTCGTGGAGGCAGGGGAGGAGTGCGACTGCGGCTCGGTGCAGGAGTGCAGCCG 1133

Qy   1573 ----tatggattatgctgtaagaaatgttccctctccaacggggctcactgcagcgacgg 1628
Db   1134 CGCAGGTGGCAACTGCTGCAAGAAATGCACCCTGACTCACGACGCCATGTGCAGCGACGG 1193

Qy   1629 gccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgccgggatgc 1688
Db   1194 GCTCTGCTGTCGCCGC------TGCAAGTACGAACCACGGGGTGTGTCCTGCCGAGAGGC 1247

Qy   1689 tgtgaacgagtgtgatattactgaatattgtactggagactctggtcagtgcccaccaaa 1748
Db   1248 CGTGAACGAGTGCGACATCGCGGAGACCTGCACCGGGGACTCTAGCCAGTGCCCGCCTAA 1307

Qy   1749 tcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctacaatggcga 1808
Db   1308 CCTGCACAAGCTGGACGGTTACTACTGTGACCATGAGCAGGGCCGCTGCTACGGAGGTCG 1367

Qy   1809 gtgcaagaccagagacaaccagtgtcagtacatctggggaacaaaggctgcagggtctga 1868
Db   1368 CTGCAAAACCCGGGACCGGCAGTGCCAGGTTCTTTGGGGCCATGCGGCTG------CTGA 1421

Qy   1869 caagttctgctatgaaaagctgaatacagaaggcactgagaagggaaactgcgggaagga 1928
Db   1422 TCGCTTCTGCTACGAGAAGCTGAATGTGGAGGGGACGGAGCGTGGGAGCTGTGGGCGCAA 1481

Qy   1929 tggagaccggtggattcagtgcagcaaacatgatgtgttctgtggattcttactctgtac 1988
Db   1482 GGGATCCGGCTGGGTCCAGTGCAGTAAGCAGGACGTGCTGTGTGGCTTCCTCCTCTGTGT 1541

Qy   1989 caatcttactcgagctccacgtattggtcaacttcagggtgagatcattccaacttcctt 2048
```

```
Db   1542 CAACATCTCTGGAGCTCCTCGGCTAGGGGACCTGGTGGGAGACATCAGTAGTGTCACCTT 1601

Qy   2049 ctaccatcaaggccgggtgattgactgcagtggtgcccatgtagttttagatgatgatac 2108
Db   1602 CTACCACCAGGGCAAGGAGCTGGACTGCAGGGGAGGCCACGTGCAGCTGGCGGACGGCTC 1661

Qy   2109 ggatgtgggctatgtagaagatggaacgccatgtggcccgtctatgatgtgtttagatcg 2168
Db   1662 TGACCTGAGCTATGTGGAGGATGGCACAGCCTGCGGGCCTAACATGTTGTGCCTGGACCA 1721

Qy   2169 gaagtgcctacaaattcaagccctaaatatgagcagctgtccactcgatttcaaggtaa 2228
Db   1722 TCGCTGCCTGCCAGCTTCTGCCTTCAACTTCAGCACCTGCCCCGGCAGTGGGGAGCGCCG 1781

     2229 agtctgttcgggccatggggtgtgtagtaatgaagccacctgcatttgtgatttcacctg 2288
     1782 GATTTGCTCCCACCACGGGGTCTGCAGCAATGAAGGGAAGTGCATCTGTCAGCCAGACTG 1841

Qy   2289 ggcagggacagattgcagtatccgggatccagt---taggaaccttcaccccccaagga 2345
Db   1842 GACAGGCAAAGACTGCAGTATCCATAACCCCCTGCCCACGTCCCCACCCACGGGGGAGAC 1901

Qy   2346 tgaaggacccaagggtcctagtgccaccaatctcataataggctccatcgctggtgccat 2405
Db   1902 GGAGAGATATAAAGGTCCCAGCGGCACCAACATCATCATTGGCTCCATCGCTGGGGCTGT 1961

Qy   2406 cctggtagcagctattgtccttgggggcacaggctggggatttaaaaatgtcaagaagag 2465
Db   1962 CCTGGTTGCAGCCTACGTCCTGGGCGGCACGGGCTGGGGATTTAAAAACATTCGCCGAGG 2021

Qy   2466 aaggttcg 2473
Db   2022 AAGGTCCG 2029


RESULT  20
I43484
LOCUS       I43484       2913 bp    DNA             PAT       07-OCT-1997
DEFINITION  Sequence 7 from patent US 5631351.
ACCESSION   I43484
VERSION     I43484.1  GI:2468728
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
  REFERENCE 1  (bases 1 to 2913)
  AUTHORS   Nakamura,Y. and Emi,M.
  TITLE     Antibodies to MDC proteins
  JOURNAL   Patent: US 5631351-A 7 20-MAY-1997;
FEATURES             Location/Qualifiers
     source          1. .2913
                     /organism="unknown"
BASE COUNT      604 a    888 c    885 g    536 t
ORIGIN


  Query Match           21.2%;  Score 530;  DB 10;  Length 2913;
  Best Local Similarity   57.8%;  Pred. No. 1.5e-116;
  Matches 1114;  Conservative    0;  Mismatches  775;  Indels   39;  Gaps    8;

Qy    573 gggtggagagcactgttactaccatggaagcatcagaggcgtcaaagactccaaggtggc 632
Db    114 GGCTGGAGACCACTGCTACTACCAGGGGAAGCTCCGGGGGAACCCGCACTCCTTCGCCGC 173

Qy    633 tctgtcaacctgcaatggacttcatggcatgtttgaagatgataccttcgtgtatatgat 692
Db    174 CCTCTCCACCTGCCAGGGGCTGCATGGGGTCTTCTCTGATGGGAACTTGACTTACATCGT 233

Qy    693 agagccactagagctggttcatgatgagaaaagcac------aggtcgaccacatataat 746
Db    234 GGAGCCCCAAGAGGTGGCTGGACCTTGGGGAGCCCCTCAGGGACCCCTTCCCCACCTCAT 293

Qy    747 ccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcactatggaaagagg 806
Db    294 TTACCGGACCCCTCTCCTCCCAGATCCCCTCGGATGCAGGGAACCAGGCTGCCTGTTTGC 353

Qy    807 tgaccagtggccctttctctctgaattacagtggttgaaaagaaggaagagagcagtgaa 866
Db    354 TGTGCCTGCCCAGTCGGCTCCTCCAAACCGGCCGAGGCTGAGAAGGAAAAGGCAGGTCCG 413

Qy    867 tccatcacgtggtatatttga---agaaatgaaatatttggaacttatgattgttaatga 923
Db    414 CCGGGGCCACCCTACAGTGCACAGTGAAACCAAGTATGTGGAGCTAATTGTGATCAACGA 473

Qy    924 tcacaaaacgtataagaagcatcgctcttctcatgcacataccaacaactttgcaaagtc 983
Db    474 CCACCAGCTGTTCGAGCAGATGCGACAGTCGGTGGTCCTCACCAGCAACTTTGCCAAGTC 533

Qy    984 cgtggtcaaccttgtggattctatttacaaggagcagctcaacaccagggttgtcctggt 1043
Db    534 CGTGGTGAACCTGGCCGATGTGATATACAAGGAGCAGCTCAACACTCGCATCGTCCTGGT 593

Qy   1044 ggctgtagagacctggactgagaaggatcagattgacatcaccaccaaccctgtgcagat 1103
Db    594 TGCCATGGAAACATGGGCAGATGGGGACAAGATCCAGGTGCAGGATGACCTCCTGGAGAC 653

Qy   1104 gctccatgagttctcaaaatac---cggcagcgcattaagcagcatgctgatgctgtgca 1160
Db    654 CCTGGCCCGGCTCATGGTCTACCGACGGGAGGGTCTGCCTGAGCCCAGTAATGCCACCCA 713

Qy   1161 cctcatctcgcgggtgacatttcactataagagaagcagtctgagttactttggaggtgt 1220
Db    714 CCTCTTCTCGGGCAGGACCTTCCAGAGCACGAGCAGCGGGGCAGCCTACGTGGGGGGCAT 773

Qy   1221 ctgttctcgcacaagaggagttggtgtgaatgagtatggtcttccaatggcagtggcaca 1280
Db    774 ATGCTCCCTGTCCCATGGCGGGGGTGTGAACGAGTACGGCAACATGGGGGCGATGGCCGT 833

Qy   1281 agtattatcgcagagcctggctcaaaaccttggaatccaatgggaaccttctagcagaaa 1340
Db    834 GACCCTTGCCCAGACGCTGGGACAGAACCTGGGCATGATGTGGAACAAACACCGGAGCTC 893

Qy   1341 gcca---aaatgtgactgcacagaatcctggggtggctgcatcatggaggaaacaggggt 1397
Db    894 GGCAGGGGACTGCAAGTGTCCAGACATCTGGCTGGGCTGCATCATGGAGGACACTGGGTT 953

Qy   1398 gtcccattctcgaaaattttcaaagtgcagcattttggagtatagagacttttacagag 1457
Db    954 CTACCTGCCCCGCAAGTTCTCTCGCTGCAGCATCGACGAGTACAACCAGTTTCTGCAGGA 1013

Qy   1458 aggaggtggagcctgccttttcaacaggccaacaaagctatttgagcccacggaatgtgg 1517
Db   1014 GGGTGGTGGCAGCTGCCTCTTCAACAAGCCCCTCAAGCTCCTGGACCCCCCAGAGTGCGG 1073

Qy   1518 aaatggatacgtggaagctggggaggagtgtgattgtggttttcatgtggaatgc----- 1572
Db   1074 GAACGGCTTCGTGGAGGCAGGGGAGGAGTGCGACTGCGGCTCGGTGCAGGAGTGCAGCCG 1133

Qy   1573 ----tatggattatgctgtaagaaaatgttccctctccaacggggctcactgcagcgacgg 1628
Db   1134 CGCAGGTGGCAACTGCTGCAAGAAATGCACCCTGACTCACGACGCCATGTGCAGCGACGG 1193

Qy   1629 gccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgccgggatgc 1688
Db   1194 GCTCTGCTGTCGCCGC------TGCAAGTACGAACCACGGGGTGTGTCCTGCCGAGAGGC 1247

Qy   1689 tgtgaacgagtgtgatattactgaatattgtactggagactctggtcagtgcccaccaaa 1748
Db   1248 CGTGAACGAGTGCGACATCGCGGAGACCTGCACCGGGGACTCTAGCCAGTGCCCGCCTAA 1307

Qy   1749 tcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctacaatggcga 1808
Db   1308 CCTGCACAAGCTGGACGGTTACTACTGTGACCATGAGCAGGGCCGCTGCTACGGAGGTCG 1367

Qy   1809 gtgcaagaccagagacaaccagtgtcagtacatctggggaacaaaggctgcagggtctga 1868
```

```
Db   1368 CTGCAAAACCCGGGACCGGCAGTGCCAGGTTCTTTGGGGCCATGCGGCTG------CTGA 1421

Qy   1869 caagttctgctatgaaaagctgaatacagaaggcactgagaagggaaactgcgggaagga 1928

Db   1422 TCGCTTCTGCTACGAGAAGCTGAATGTGGAGGGGACGGAGCGTGGGAGCTGTGGGCGCAA 1481

Qy   1929 tggagaccggtggattcagtgcagcaaacatgatgtgttctgtggattcttactctgtac 1988

Db   1482 GGGATCCGGCTGGGTCCAGTGCAGTAAGCAGGACGTGCTGTGTGGCTTCCTCCTCTGTGT 1541

Qy   1989 caatcttactcgagctccacgtattggtcaacttcagggtgagatcattccaacttcctt 2048

Db   1542 CAACATCTCTGGAGCTCCTCGGCTAGGGGACCTGGTGGGAGACATCAGTAGTGTCACCTT 1601

Qy   2049 ctaccatcaaggccgggtgattgactgcagtggtgcccatgtagttttagatgatgatac 2108

Db   1602 CTACCACCAGGGCAAGGAGCTGGACTGCAGGGGAGGCCACGTGCAGCTGGCGGACGGCTC 1661

Qy   2109 ggatgtgggctatgtagaagatggaacgccatgtggcccgtctatgatgtgtttagatcg 2168

Db   1662 TGACCTGAGCTATGTGGAGGATGGCACAGCCTGCGGGCCTAACATGTTGTGCCTGGACCA 1721

     2169 gaagtgcctacaaattcaagccctaaatatgagcagctgtccactcgattccaagggtaa 2228

Db   1722 TCGCTGCCTGCCAGCTTCTGCCTTCAACTTCAGCACCTGCCCCGGCAGTGGGGAGCGCCG 1781

Qy   2229 agtctgttcgggccatggggtgtgtagtaatgaagccacctgcatttgtgatttcacctg 2288

Db   1782 GATTTGCTCCCACCACGGGGTCTGCAGCAATGAAGGGAAGTGCATCTGTCAGCCAGACTG 1841

Qy   2289 ggcagggacagattgcagtatccgggatccagt---taggaaccttcacccccccaagga 2345

Db   1842 GACAGGCAAAGACTGCAGTATCCATAACCCCCTGCCCACGTCCCCACCCACGGGGGAGAC 1901

Qy   2346 tgaaggacccaagggtcctagtgccaccaatctcataataggctccatcgctggtgccat 2405

Db   1902 GGAGAGATATAAAGGTCCCAGCGGCACCAACATCATCATTGGCTCCATCGCTGGGGCTGT 1961

Qy   2406 cctggtagcagctattgtccttgggggcacaggctggggatttaaaaatgtcaagaagag 2465

Db   1962 CCTGGTTGCAGCCTACGTCCTGGGCGGCACGGGCTGGGGATTTAAAAACATTCGCCGAGG 2021

Qy   2466 aaggttcg 2473

Db   2022 AAGGTCCG 2029


Search completed: August 15, 2001, 21:52:25
Job time: 7788 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2001, 19:46:52 ; Search time 81.94 Seconds
(without alignments)
5773.597 Million cell updates/sec

Title: US-09-634-252A-2
Perfect score: 2499
Sequence: 1 atgaagccgcccggcagcag.....ctcagcaaggcccatctga 2499

[redacted]ing table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 324599 seqs, 94655562 residues

Total number of hits satisfying chosen parameters: 649198

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 60 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	575	23.0	3183	1	US-08-243-542-8	Sequence 8, Appli
2	575	23.0	3183	1	US-08-477-407-8	Sequence 8, Appli
3	575	23.0	3183	1	US-08-484-355-8	Sequence 8, Appli
4	530	21.2	2913	1	US-08-243-542-7	Sequence 7, Appli
5	530	21.2	2913	1	US-08-477-407-7	Sequence 7, Appli
6	530	21.2	2913	1	US-08-484-355-7	Sequence 7, Appli
7	513.2	20.5	2923	1	US-08-243-542-6	Sequence 6, Appli
8	513.2	20.5	2923	1	US-08-477-407-6	Sequence 6, Appli
9	513.2	20.5	2923	1	US-08-484-355-6	Sequence 6, Appli
10	338.4	13.5	1464	1	US-08-243-542-5	Sequence 5, Appli
11	338.4	13.5	1464	1	US-08-477-407-5	Sequence 5, Appli
12	338.4	13.5	1464	1	US-08-484-355-5	Sequence 5, Appli
13	107.6	4.3	2648	2	US-08-836-443-1	Sequence 1, Appli
14	99.6	4.0	2251	2	US-08-836-443-2	Sequence 2, Appli
15	71.8	2.9	1463	5	PCT-US93-05640-10	Sequence 10, Appl
16	66.6	2.7	9278	1	US-08-243-542-9	Sequence 9, Appli
17	66.6	2.7	9278	1	US-08-477-407-9	Sequence 9, Appli
18	66.6	2.7	9278	1	US-08-484-355-9	Sequence 9, Appli
19	64	2.6	2373	1	US-08-264-101-1	Sequence 1, Appli
20	64	2.6	2373	2	US-08-765-243-1	Sequence 1, Appli
21	64	2.6	2373	5	PCT-US95-07295-1	Sequence 1, Appli
22	64	2.6	2650	2	US-08-765-243-7	Sequence 7, Appli
23	64	2.6	2650	5	PCT-US95-07295-7	Sequence 7, Appli
24	50.8	2.0	152331	3	US-09-128-155-16	Sequence 16, Appl
c 25	48.2	1.9	3978	3	US-08-726-214-1	Sequence 1, Appli
26	46.6	1.9	7218	1	US-08-232-463-14	Sequence 14, Appl
c 27	46.4	1.9	320	4	US-09-165-264-7	Sequence 7, Appli
c 28	46.4	1.9	7218	1	US-08-232-463-14	Sequence 14, Appl
c 29	46.2	1.8	319	4	US-09-165-264-8	Sequence 8, Appli
c 30	45.8	1.8	318	4	US-09-165-264-12	Sequence 12, Appl
c 31	45.8	1.8	320	4	US-09-165-264-11	Sequence 11, Appl
32	45.6	1.8	3181	3	US-09-135-021-1	Sequence 1, Appli
33	45.6	1.8	3182	3	US-09-135-021-5	Sequence 5, Appli
34	45.4	1.8	12001	1	US-08-458-568A-11	Sequence 11, Appl
35	45	1.8	3177	1	US-08-042-747A-4	Sequence 4, Appli
c 36	44.8	1.8	320	4	US-09-165-264-14	Sequence 14, Appl
c 37	44.4	1.8	320	4	US-09-165-264-13	Sequence 13, Appl
c 38	44.2	1.8	1028	4	US-08-118-200-1	Sequence 1, Appli
c 39	44.2	1.8	1028	4	US-08-458-745-1	Sequence 1, Appli
40	44.2	1.8	50937	4	US-09-428-517-1	Sequence 1, Appli
41	44	1.8	1768	1	US-08-264-101-3	Sequence 3, Appli
42	44	1.8	1768	2	US-08-765-243-3	Sequence 3, Appli
43	44	1.8	1768	5	PCT-US95-07295-3	Sequence 3, Appli
44	44	1.8	2553	2	US-08-765-243-5	Sequence 5, Appli
45	44	1.8	2553	5	PCT-US95-07295-5	Sequence 5, Appli
46	43.6	1.7	503	4	US-08-483-533-7	Sequence 7, Appli
47	43.6	1.7	1292	4	US-08-483-533-37	Sequence 37, Appl
48	43.2	1.7	2353	5	PCT-US92-06840-1	Sequence 1, Appli
49	42.8	1.7	4257	2	US-08-690-473-1	Sequence 1, Appli
50	42.8	1.7	4257	4	US-09-259-821A-1	Sequence 1, Appli
51	42.8	1.7	4257	4	US-08-843-659-1	Sequence 1, Appli
c 52	42.8	1.7	12001	1	US-08-458-568A-11	Sequence 11, Appl
c 53	42.8	1.7	18994	1	US-08-459-586-4	Sequence 4, Appli
c 54	42.8	1.7	18994	2	US-08-282-696-4	Sequence 4, Appli
c 55	42.6	1.7	4060	1	US-08-308-949A-1	Sequence 1, Appli
56	42	1.7	1590	2	US-08-836-442-2	Sequence 2, Appli
57	42	1.7	2056	2	US-08-836-442-1	Sequence 1, Appli
c 58	42	1.7	3647	1	US-07-914-281-7	Sequence 7, Appli
c 59	42	1.7	3647	1	US-08-393-246-7	Sequence 7, Appli
c 60	42	1.7	3647	1	US-08-525-058A-7	Sequence 7, Appli

ALIGNMENTS

```
RESULT 1
US-08-243-542-8
; Sequence 8, Application US/08243542
; Patent No. 5552526
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  : EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/243,542
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
```

```
;       APPLICATION NUMBER: JP 6-84470
;       FILING DATE:  22 APRIL 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Terryence F. Chapman
;       REGISTRATION NUMBER:  32 549
;       REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (616) 381-1156
;       TELEFAX:  (616) 381-5465
;   INFORMATION FOR SEQ ID NO:  8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  3183 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA to mRNA
;     ORIGINAL SOURCE:
;       ORGANISM:  Homo sapiens
;     IMMEDIATE SOURCE:
;       LIBRARY:  human fetal brain cDNA library
      FEATURE:
        NAME/KEY:  3' UTR
        LOCATION:  2308..3183
      FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..2307
US-08-243-542-8

  Query Match             23.0%;  Score 575;  DB 1;  Length 3183;
  Best Local Similarity   57.5%;  Pred. No. 2.4e-151;
  Matches 1226;  Conservative    0;  Mismatches  860;  Indels   45;  Gaps    9;

Qy     376 gaaagcccttatcacgttcttgacacaaaggcaagacaccagcaaaaacataataaggct 435
Db     181 GAGGTCCGAAAGCAGCAGCTGGACACAAGGGTCCGCCAGGAGCCACCAGGGGGCCCGCCT 240

Qy     436 gtccatctggcccaggcaagcttccagattgaagccttcggctccaaattcattcttgac 495
Db     241 GTCCATCTGGCCCAGGTGAGTTTCGTCATCCCAGCCTTCAACTCAAACTTCACCCTGGAC 300

Qy     496 ctcatactgaacaatggtttgttgtcttctgattatgtggagattcactac------gaa 549
Db     301 CTGGAGCTGAACCACCACCTCCTCTCCTCGCAATACGTGGAGCGCCACTTCAGCCGGGAG 360

Qy     550 aatgggaaaccacagtactctaagggtggagagcactgttactaccatggaagcatcaga 609
Db     361 GGGACAACCCAGCACAGCACCGGGGCTGGAGACCACTGCTACTACCAGGGGAAGCTCCGG 420

Qy     610 ggcgtcaaagactccaaggtggctctgtcaacctgcaatggacttcatggcatgtttgaa 669
Db     421 GGGAACCCGCACTCCTTCGCCGCCCTCTCCACCTGCCAGGGGCTGCATGGGGTCTTCTCT 480

Qy     670 gatgataccttcgtgtatatgatagagccactagagctggttcatgatgagaaaagcac- 728
Db     481 GATGGGAACTTGACTTACATCGTGGAGCCCCAAGAGGTGGCTGGACCTTGGGGAGCCCCT 540

Qy     729 -----aggtcgaccacatataatccagaaaaccttggcaggacagtattctaagcaaatg 783
Db     541 CAGGGACCCCTTCCCCACCTCATTTACCGGACCCCTCTCCTCCCAGATCCCCTCGGATGC 600

Qy     784 aagaatctcactatggaaagaggtgaccagtggccctttctctctgaattacagtggttg 843
Db     601 AGGGAACCAGGCTGCCTGTTTGCTGTGCCTGCCCAGTCGGCTCCTCCAAACCGGCCGAGG 660

Qy     844 aaaagaaggaagagagcagtgaatccatcacgtggtatatttga---agaaatgaaatat 900
Db     661 CTGAGAAGGAAAAGGCAGGTCCGCCGGGGCCACCCTACAGTGCACAGTGAAACCAAGTAT 720

Qy     901 ttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgca 960
Db     721 GTGGAGCTAATTGTGATCAACGACCACCAGCTGTTCGAGCAGATGCGACAGTCGGTGGTC 780

Qy     961 cataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcag 1020
Db     781 CTCACCAGCAACTTTGCCAAGTCCGTGGTGAACCTGGCCGATGTGATATACAAGGAGCAG 840

Qy    1021 ctcaacaccagggttgtcctggtggctgtagagacctggactgagaaggatcagattgac 1080
Db     841 CTCAACACTCGCATCGTCCTGGTTGCCATGGAAACATGGGCAGATGGGGACAAGATCCAG 900

Qy    1081 atcaccaccaaccctgtgcagatgctccatgagttctcaaaatac---cggcagcgcatt 1137
Db     901 GTGCAGGATGACCTCCTGGAGACCCTGGCCCGGCTCATGGTCTACCGACGGGAGGGTCTG 960

Qy    1138 aagcagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagc 1197
Db     961 CCTGAGCCCAGTAATGCCACCCACCTCTTCTCGGGCAGGACCTTCCAGAGCACGAGCAGC 1020

Qy    1198 agtctgagttactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtat 1257
Db    1021 GGGGCAGCCTACGTGGGGGGCATATGCTCCCTGTCCCATGGCGGGGGTGTGAACGAGTAC 1080

Qy    1258 ggtcttccaatggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatc 1317
Db    1081 GGCAACATGGGGGCGATGGCCGTGACCCTTGCCCAGACGCTGGGACAGAACCTGGGCATG 1140

Qy    1318 caatgggaaccttctagcagaaagccaa---aatgtgactgcacagaatcctggggtggc 1374
Db    1141 ATGTGGAACAAACACCGGAGCTCGGCAGGGGACTGCAAGTGTCCAGACATCTGGCTGGGC 1200

Qy    1375 tgcatcatggaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttg 1434
Db    1201 TGCATCATGGAGGACACTGGGTTCTACCTGCCCCGCAAGTTCTCTCGCTGCAGCATCGAC 1260

Qy    1435 gagtatagagacttttttacagagaggaggtggagcctgccttttcaacaggccaacaaag 1494
Db    1261 GAGTACAACCAGTTTCTGCAGGAGGGTGGTGGCAGCTGCCTCTTCAACAAGCCCCTCAAG 1320

Qy    1495 ctatttgagcccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgt 1554
Db    1321 CTCCTGGACCCCCCAGAGTGCGGGAACGGCTTCGTGGAGGCAGGGGAGGAGTGCGACTGC 1380

Qy    1555 ggttttcatgtggaatgc---------tatggattatgctgtaagaaatgttccctctcc 1605
Db    1381 GGCTCGGTGCAGGAGTGCAGCCGCGCAGGTGGCAACTGCTGCAAGAAATGCACCCTGACT 1440

Qy    1606 aacggggctcactgcagcgacgggccctgctgtaacaatacctcatgtctttttcagcca 1665
Db    1441 CACGACGCCATGTGCAGCGACGGGCTCTGCTG------TCGCCGCTGCAAGTACGAACCA 1494

Qy    1666 cgagggtatgaatgccgggatgctgtgaacgagtgtgatattactgaatattgtactgga 1725
Db    1495 CGGGGTGTGTCCTGCCGAGAGGCCGTGAACGAGTGCGACATCGCGGAGACCTGCACCGGG 1554

Qy    1726 gactctggtcagtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaat 1785
Db    1555 GACTCTAGCCAGTGCCCGCCTAACCTGCACAAGCTGGACGGTTACTACTGTGACCATGAG 1614

Qy    1786 cagggccgctgctacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctgg 1845
Db    1615 CAGGGCCGCTGCTACGGAGGTCGCTGCAAAACCCGGGACCGGCAGTGCCAGGTTCTTTGG 1674

Qy    1846 ggaacaaaggctgcagggtctgacaagttctgctatgaaaagctgaatacagaaggcact 1905
Db    1675 GGCCATGCGGCTG------CTGATCGCTTCTGCTACGAGAAGCTGAATGTGGAGGGGACG 1728

Qy    1906 gagaagggaaactgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtg 1965
Db    1729 GAGCGTGGGAGCTGTGGGCGCAAGGGATCCGGCTGGGTCCAGTGCAGTAAGCAGGACGTG 1788

Qy    1966 ttctgtggattcttactctgtaccaatcttactcgagctccacgtattggtcaacttcag 2025
Db    1789 CTGTGTGGCTTCCTCCTCTGTGTCAACATCTCTGGAGCTCCTCGGCTAGGGGACCTGGTG 1848

Qy    2026 ggtgagatcattccaacttccttctaccatcaaggccgggtgattgactgcagtggtgcc 2085
```

```
Db   1849 GGAGACATCAGTAGTGTCACCTTCTACCACCAGGGCAAGGAGCTGGACTGCAGGGGAGGC 1908

Qy   2086 catgtagttttagatgatgatacggatgtgggctatgtagaagatggaacgccatgtggc 2145
Db   1909 CACGTGCAGCTGGCGGACGGCTCTGACCTGAGCTATGTGGAGGATGGCACAGCCTGCGGG 1968

Qy   2146 ccgtctatgatgtgtttagatcggaagtgcctacaaattcaagccctaaatatgagcagc 2205
Db   1969 CCTAACATGTTGTGCCTGGACCATCGCTGCCTGCCAGCTTCTGCCTTCAACTTCAGCACC 2028

Qy   2206 tgtccactcgattccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagcc 2265
Db   2029 TGCCCCGGCAGTGGGGAGCGCCGGATTTGCTCCCACCACGGGGTCTGCAGCAATGAAGGG 2088

Qy   2266 acctgcatttgtgatttcacctgggcagggacagattgcagtatccgggatccagttagg 2325
Db   2089 AAGTGCATCTGTCAGCCAGACTGGACAGGCAAAGACTGCAGTATCCATAACCCCCTGCCC 2148

Qy   2326 aaccttcacccccccaaggat---gaaggacccaagggtcctagtgccaccaatctcata 2382
Db   2149 ACGTCCCCACCCACGGGGGAGACGGAGAGATATAAAGGTCCCAGCGGCACCAACATCATC 2208

Qy   2383 ataggctccatcgctggtgccatcctggtagcagctattgtccttgggggcacaggctgg 2442
Db   2209 ATTGGCTCCATCGCTGGGGCTGTCCTGGTTGCAGCCATCGTCCTGGGCGGCACGGGCTGG 2268

Qy   2443 ggatttaaaaatgtcaagaagagaaggttcg 2473
Db   2269 GGATTTAAAAACATTCGCCGAGGAAGGTCCG 2299


RESULT   2
US-08-477-407-8
; Sequence 8, Application US/08477407
; Patent No. 5631351
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  :  EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,407
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/243,542
;      FILING DATE:  13-MAY-1994
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
;      TELEFAX:  (616) 381-5465
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3183 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY:  human fetal brain cDNA library
;    FEATURE:
;      NAME/KEY:  3' UTR
;      LOCATION:  2308..3183
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2307
US-08-477-407-8


  Query Match             23.0%;  Score 575;  DB 1;  Length 3183;
  Best Local Similarity   57.5%;  Pred. No. 2.4e-151;
  Matches 1226;  Conservative    0;  Mismatches  860;  Indels   45;  Gaps    9;

Qy    376 gaaagcccttatcacgttcttgacacaaaggcaagacaccagcaaaaacataataaggct 435
Db    181 GAGGTCCGAAAGCAGCAGCTGGACACAAGGGTCCGCCAGGAGCCACCAGGGGGCCCGCCT 240

Qy    436 gtccatctggcccaggcaagcttccagattgaagccttcggctccaaattcattcttgac 495
Db    241 GTCCATCTGGCCCAGGTGAGTTTCGTCATCCCAGCCTTCAACTCAAACTTCACCCTGGAC 300

Qy    496 ctcatactgaacaatggtttgttgtcttctgattatgtggagattcactac------gaa 549
Db    301 CTGGAGCTGAACCACCACCTCCTCTCCTCGCAATACGTGGAGCGCCACTTCAGCCGGGAG 360

Qy    550 aatgggaaaccacagtactctaagggtggagagcactgttactaccatggaagcatcaga 609
Db    361 GGGACAACCCAGCACAGCACCGGGGCTGGAGACCACTGCTACTACCAGGGGAAGCTCCGG 420

Qy    610 ggcgtcaaagactccaaggtggctctgtcaacctgcaatggacttcatggcatgtttgaa 669
Db    421 GGGAACCCGCACTCCTTCGCCGCCCTCTCCACCTGCCAGGGGCTGCATGGGGTCTTCTCT 480

Qy    670 gatgataccttcgtgtatatgatagagccactagagctggttcatgatgagaaaagcac- 728
Db    481 GATGGGAACTTGACTTACATCGTGGAGCCCCAAGAGGTGGCTGGACCTTGGGGAGCCCCT 540

Qy    729 -----aggtcgaccacatataatccagaaaaccttggcaggacagtattctaagcaaatg 783
Db    541 CAGGGACCCCTTCCCCACCTCATTTACCGGACCCCTCTCCTCCCAGATCCCCTCGGATGC 600

Qy    784 aagaatctcactatggaaagaggtgaccagtggccctttctctctgaattacagtggttg 843
Db    601 AGGGAACCAGGCTGCCTGTTTGCTGTGCCTGCCCAGTCGGCTCCTCCAAACCGGCCGAGG 660

Qy    844 aaaagaaggaagagagcagtgaatccatcacgtggtatatttga---agaaatgaaatat 900
Db    661 CTGAGAAGGAAAAGGCAGGTCCGCCGGGGCCACCCTACAGTGCACAGTGAAACCAAGTAT 720

Qy    901 ttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgca 960
Db    721 GTGGAGCTAATTGTGATCAACGACCACCAGCTGTTCGAGCAGATGCGACAGTCGGTGGTC 780

Qy    961 cataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttacaaggagcag 1020
Db    781 CTCACCAGCAACTTTGCCAAGTCCGTGGTGAACCTGGCCGATGTGATATACAAGGAGCAG 840
```

```
Qy   1021 ctcaacaccagggttgtcctggtggctgtagagacctggactgagaaggatcagattgac 1080
Db    841 CTCAACACTCGCATCGTCCTGGTTGCCATGGAAACATGGGCAGATGGGGACAAGATCCAG 900

Qy   1081 atcaccaccaaccctgtgcagatgctccatgagttctcaaaatac---cggcagcgcatt 1137
Db    901 GTGCAGGATGACCTCCTGGAGACCCTGGCCCGGCTCATGGTCTACCGACGGGAGGGTCTG 960

Qy   1138 aagcagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagc 1197
Db    961 CCTGAGCCCAGTAATGCCACCCACCTCTTCTCGGGCAGGACCTTCCAGAGCACGAGCAGC 1020

Qy   1198 agtctgagttactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtat 1257
Db   1021 GGGGCAGCCTACGTGGGGGGCATATGCTCCCTGTCCCATGGCGGGGGTGTGAACGAGTAC 1080

Qy   1258 ggtcttccaatggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatc 1317
Db   1081 GGCAACATGGGGGCGATGGCCGTGACCCTTGCCCAGACGCTGGGACAGAACCTGGGCATG 1140

Qy   1318 caatgggaaccttctagcagaaagccaa---aatgtgactgcacagaatcctggggtggc 1374
Db   1141 ATGTGGAACAAACACCGGAGCTCGGCAGGGGACTGCAAGTGTCCAGACATCTGGCTGGGC 1200

Qy   1375 tgcatcatggaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttg 1434
Db   1201 TGCATCATGGAGGACACTGGGTTCTACCTGCCCCGCAAGTTCTCTCGCTGCAGCATCGAC 1260

Qy   1435 gagtatagagacttttttacagagaggaggtggagcctgccttttcaacaggccaacaaag 1494
Db   1261 GAGTACAACCAGTTTCTGCAGGAGGGTGGTGGCAGCTGCCTCTTCAACAAGCCCCTCAAG 1320

Qy   1495 ctatttgagcccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgt 1554
Db   1321 CTCCTGGACCCCCCAGAGTGCGGGAACGGCTTCGTGGAGGCAGGGGAGGAGTGCGACTGC 1380

Qy   1555 ggttttcatgtggaatgc---------tatggattatgctgtaagaaatgttccctctcc 1605
Db   1381 GGCTCGGTGCAGGAGTGCAGCCGCGCAGGTGGCAACTGCTGCAAGAAATGCACCCTGACT 1440

Qy   1606 aacggggctcactgcagcgacgggccctgctgtaacaatacctcatgtctttttcagcca 1665
Db   1441 CACGACGCCATGTGCAGCGACGGGCTCTGCTG------TCGCCGCTGCAAGTACGAACCA 1494

Qy   1666 cgagggtatgaatgccgggatgctgtgaacgagtgtgatattactgaatattgtactgga 1725
Db   1495 CGGGGTGTGTCCTGCCGAGAGGCCGTGAACGAGTGCGACATCGCGGAGACCTGCACCGGG 1554

Qy   1726 gactctggtcagtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaat 1785
Db   1555 GACTCTAGCCAGTGCCCGCCTAACCTGCACAAGCTGGACGGTTACTACTGTGACCATGAG 1614

     1786 cagggccgctgctacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctgg 1845
     1615 CAGGGCCGCTGCTACGGAGGTCGCTGCAAAACCCGGGACCGGCAGTGCCAGGTTCTTTGG 1674

Qy   1846 ggaacaaaggctgcagggtctgacaagttctgctatgaaaagctgaatacagaaggcact 1905
Db   1675 GGCCATGCGGCTG------CTGATCGCTTCTGCTACGAGAAGCTGAATGTGGAGGGGACG 1728

Qy   1906 gagaagggaaactgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtg 1965
Db   1729 GAGCGTGGGAGCTGTGGGCGCAAGGGATCCGGCTGGGTCCAGTGCAGTAAGCAGGACGTG 1788

Qy   1966 ttctgtggattcttactctgtaccaatcttactcgagctccacgtattggtcaacttcag 2025
Db   1789 CTGTGTGGCTTCCTCCTCTGTGTCAACATCTCTGGAGCTCCTCGGCTAGGGGACCTGGTG 1848

Qy   2026 ggtgagatcattccaacttccttctaccatcaaggccgggtgattgactgcagtggtgcc 2085
Db   1849 GGAGACATCAGTAGTGTCACCTTCTACCACCAGGGCAAGGAGCTGGACTGCAGGGGAGGC 1908

Qy   2086 catgtagttttagatgatgatacggatgtgggctatgtagaagatggaacgccatgtggc 2145
Db   1909 CACGTGCAGCTGGCGGACGGCTCTGACCTGAGCTATGTGGAGGATGGCACAGCCTGCGGG 1968

Qy   2146 ccgtctatgatgtgtttagatcggaagtgcctacaaattcaagccctaaatatgagcagc 2205
Db   1969 CCTAACATGTTGTGCCTGGACCATCGCTGCCTGCCAGCTTCTGCCTTCAACTTCAGCACC 2028

Qy   2206 tgtccactcgattccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagcc 2265
Db   2029 TGCCCCGGCAGTGGGGAGCGCCGGATTTGCTCCCACCACGGGGTCTGCAGCAATGAAGGG 2088

Qy   2266 acctgcatttgtgatttcacctgggcagggacagattgcagtatccgggatccagttagg 2325
Db   2089 AAGTGCATCTGTCAGCCAGACTGGACAGGCAAAGACTGCAGTATCCATAACCCCCTGCCC 2148

Qy   2326 aaccttcaccccccccaaggat---gaaggacccaagggtcctagtgccaccaatctcata 2382
Db   2149 ACGTCCCCACCCACGGGGGAGACGGAGAGATATAAAGGTCCCAGCGGCACCAACATCATC 2208

Qy   2383 ataggctccatcgctggtgccatcctggtagcagctattgtccttgggggcacaggctgg 2442
Db   2209 ATTGGCTCCATCGCTGGGGCTGTCCTGGTTGCAGCCATCGTCCTGGGCGGCACGGGCTGG 2268

Qy   2443 ggatttaaaaaatgtcaagaagagaaggttcg 2473
Db   2269 GGATTTAAAAACATTCGCCGAGGAAGGTCCG 2299
```

```
RESULT   3
US-08-484-355-8
; Sequence 8, Application US/08484355
; Patent No. 5705341
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  :  EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,355
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/243,542
;      FILING DATE:  13-MAY-1994
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
```

```
;      TELEFAX:  (616) 381-5465
;  INFORMATION FOR SEQ ID NO:  8:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3183 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY:  human fetal brain cDNA library
;    FEATURE:
;      NAME/KEY:  3' UTR
;      LOCATION:  2308..3183
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..2307
US-08-484-355-8


  Query Match             23.0%;  Score 575;  DB 1;  Length 3183;
  Best Local Similarity   57.5%;  Pred. No. 2.4e-151;
  Matches 1226;  Conservative    0;  Mismatches  860;  Indels   45;  Gaps    9;

Qy    376 gaaagcccttatcacgttcttgacacaaaggcaagacaccagcaaaaacataataaggct 435
          ||   ||   | ||    || |||||| ||   | ||  ||| |  |        | ||
Db    181 GAGGTCCGAAAGCAGCAGCTGGACACAAGGGTCCGCCAGGAGCCACCAGGGGGCCCGCCT 240

Qy    436 gtccatctggcccaggcaagcttccagattgaagccttcggctccaaattcattcttgac 495
          ||||||||||||||||  || |||   ||   |||||||  ||| || ||||  || |||
Db    241 GTCCATCTGGCCCAGGTGAGTTTCGTCATCCCAGCCTTCAACTCAAACTTCACCCTGGAC 300

Qy    496 ctcatactgaacaatggtttgttgtcttctgattatgtggagattcactac------gaa 549
          ||    ||||| |      |  || ||  | || ||||||   |||| |       ||
Db    301 CTGGAGCTGAACCACCACCTCCTCTCCTCGCAATACGTGGAGCGCCACTTCAGCCGGGAG 360

Qy    550 aatgggaaaccacagtactctaagggtggagagcactgttactaccatggaagcatcaga 609
              |  |  ||   | |   || |||||| ||||| |||||||| || |   || |
Db    361 GGGACAACCCAGCACAGCACCGGGGCTGGAGACCACTGCTACTACCAGGGGAAGCTCCGG 420

Qy    610 ggcgtcaaagactccaaggtggctctgtcaacctgcaatggacttcatggcatgtttgaa 669
          ||   |  |||||   |  || || ||  |||||| | || ||  |||||  | ||
Db    421 GGGAACCCGCACTCCTTCGCCGCCCTCTCCACCTGCCAGGGGCTGCATGGGGTCTTCTCT 480

      670 gatgataccttcgtgtatatgatagagccactagagctggttcatgatgagaaaagcac- 728
          ||||  | |||    || ||  | ||||| | |||| |||. |    |  |  |  | |
      481 GATGGGAACTTGACTTACATCGTGGAGCCCCAAGAGGTGGCTGGACCTTGGGGAGCCCCT 540

Qy    729 -----aggtcgaccacatataatccagaaaaccttggcaggacagtattctaagcaaatg 783
               |   | || ||  | ||  |    |||         |   || |      |
Db    541 CAGGGACCCCTTCCCCACCTCATTTACCGGACCCCTCTCCTCCCAGATCCCCTCGGATGC 600

Qy    784 aagaatctcactatggaaagaggtgaccagtggccctttctctctgaattacagtggttg 843
          | | | |              | ||  |    |  |      ||  |   | |  |  |
Db    601 AGGGAACCAGGCTGCCTGTTTGCTGTGCCTGCCCAGTCGGCTCCTCCAAACCGGCCGAGG 660

Qy    844 aaaagaaggaagagagcagtgaatccatcacgtggtatatttga---agaaatgaaatat 900
              |||||||| ||    ||  |      |    || | |  |   .||||  || |||
Db    661 CTGAGAAGGAAAAGGCAGGTCCGCCGGGGCCACCCTACAGTGCACACTGAAACCAAGTAT 720

Qy    901 ttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctcttctcatgca 960
           |||| || ||  |  | || || ||| |   ||   || ||   ||    ||     |
Db    721 GTGGAGCTAATTGTGATCAACGACCACCAGCTGTTCGAGCAGATGCCACAGTCGGTGGTC 780

Qy    961 cataccaacaactttgcaaagtccgtggtcaaccttgtggattctattacaaggagcag 1020
          |  |||| |||||||| |||||||||| ||||| |  |||   || ||||||||||||
Db    781 CTCACCAGCAACTTTGCCAAGTCCGTGGTGAACCTGGCCGATGTGATATACAAGGAGCAG 840

Qy   1021 ctcaacaccagggttgtcctggtggctgtagagacctggactgagaaggatcagattgac 1080
          ||||||||  |  | |||||||| ||  | || || ||| | ||   |||  ||||  |
Db    841 CTCAACACTCGCATCGTCCTGGTTGCCATGGAAACATGGGCAGATGCGGACAAGATCCAG 900

Qy   1081 atcaccaccaaccctgtgcagatgctccatgagttctcaaaatac---cggcagcgcatt 1137
           |        |||   || |||  ||      | ||      |||   ||| || |  |
Db    901 GTGCAGGATGACCTCCTGGAGACCCTGGCCCGGCTCATGGTCTACCGACGGGAGGGTCTG 960

Qy   1138 aagcagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactataagagaagc 1197
             |||    | ||||   |||||| ||||| |   ||| || ||    | ||| |||
Db    961 CCTGAGCCCAGTAATGCCACCCACCTCTTCTCGGGCAGGACCTTCCAGAGCACGAGCAGC 1020

Qy   1198 agtctgagttactttggaggtgtctgttctcgcacaagaggagttggtgtgaatgagtat 1257
           |      ||| | || ||  | || || |    |    || |  |||||||| ||||
Db   1021 GGGGCAGCCTACGTGGGGGGCATATGCTCCCTGTCCCATGGCGGGGGTGTGAACGAGTAC 1080

Qy   1258 ggtcttccaatggcagtggcacaagtattatcgcagagcctggctcaaaaccttggaatc 1317
           ||      |||  ||||        |  | ||||  ||||  || ||||| || ||
Db   1081 GGCAACATGGGGGCGATGGCCGTGACCCTTGCCCAGACGCTGGGACAGAACCTGGGCATG 1140

Qy   1318 caatgggaaccttctagcagaaagccaa---aatgtgactgcacagaatcctggggtggc 1374
             ||| |       | ||   | ||     | ||  | ||  ||||   ||||   |||
Db   1141 ATGTGGAACAAACACCGGAGCTCGGCAGGGGACTGCAAGTGTCCAGACATCTGGCTGGGC 1200

Qy   1375 tgcatcatggaggaaacaggggtgtcccattctcgaaaattttcaaagtgcagcattttg 1434
          ||||||||||||||| || ||| | | ||   | || || || ||    ||||||||
Db   1201 TGCATCATGGAGGACACTGGGTTCTACCTGCCCCGCAAGTTCTCTCGCTGCAGCATCGAC 1260

Qy   1435 gagtatagagacttttacagagaggaggtggagcctgccttttcaacaggccaacaaag 1494
          ||||| |   | ||| | |||   || |||||  |||||| ||||||| |||   |||
Db   1261 GAGTACAACCAGTTTCTGCAGGAGGGTGGTGGCAGCTGCCTCTTCAACAAGCCCCTCAAG 1320

Qy   1495 ctatttgagcccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgt 1554
          ||  | || ||| | || || || || || | |||||| || |||||||||||| || ||
Db   1321 CTCCTGGACCCCCCAGAGTGCGGGAACGGCTTCGTGGAGGCAGGGGAGGAGTGCGACTGC 1380

Qy   1555 ggttttcatgtggaatgc---------tatggattatgctgtaagaaatgttccctctcc 1605
          || |      ||| |||           |||   ||||| |||||||  ||||  |
Db   1381 GGCTCGGTGCAGGAGTGCAGCCGCGCAGGTGGCAACTGCTGCAAGAAATGCACCCTGACT 1440

Qy   1606 aacggggctcactgcagcgacgggccctgctgtaacaatacctcatgtctttttcagcca 1665
           ||| ||    |||||||||||||| ||||||      |  |   ||    |  | |||
Db   1441 CACGACGCCATGTGCAGCGACGGGCTCTGCTG------TCGCCGCTGCAAGTACGAACCA 1494

Qy   1666 cgagggtatgaatgccgggatgctgtgaacgagtgtgatattactgaatattgtactgga 1725
          || ||       ||||| || || ||||||||||| || ||  | ||    || || ||
Db   1495 CGGGGTGTGTCCTGCCGAGAGGCCGTGAACGAGTGCGACATCGCGGAGACCTGCACCGGG 1554

Qy   1726 gactctggtcagtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaat 1785
          |||||| | |||||||| || || || || ||||  |||||  ||    ||  | ||  |
Db   1555 GACTCTAGCCAGTGCCCGCCTAACCTGCACAAGCTGGACGGTTACTACTGTGACCATGAG 1614

Qy   1786 cagggccgctgctacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctgg 1845
          ||||||||||||||||  ||    ||||| ||| | |||   |||||  |||    | |||
Db   1615 CAGGGCCGCTGCTACGGAGGTCGCTGCAAAACCCGGGACCGGCAGTGCCAGGTTCTTTGG 1674

Qy   1846 ggaacaaaggctgcagggtctgacaagttctgctatgaaaagctgaatacagaaggcact 1905
          ||      |||||      ||||    |||||||| || |||||||||   || || ||
Db   1675 GGCCATGCGGCTG------CTGATCGCTTCTGCTACGAGAAGCTGAATGTGGAGGGGACG 1728

Qy   1906 gagaagggaaactgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtg 1965
          |||   || | ||| |||    | |||  | | ||| | |||||||| || || || |||
Db   1729 GAGCGTGGGAGCTGTGGGCGCAAGGGATCCGGCTGGGTCCAGTGCAGTAAGCAGGACGTG 1788

Qy   1966 ttctgtggattcttactctgtaccaatcttactcgagctccacgtattggtcaacttcag 2025
           | ||||| ||| | |||||  ||| |  || |||||||  ||  | ||  | ||   |
Db   1789 CTGTGTGGCTTCCTCCTCTGTGTCAACATCTCTGGAGCTCCTCGGCTAGGGGACCTGGTG 1848

Qy   2026 ggtgagatcattccaacttccttctaccatcaaggccgggtgattgactgcagtggtgcc 2085
          || || |||| |       |||||||||| || |||  || | | ||||||| || | |
Db   1849 GGAGACATCAGTAGTGTCACCTTCTACCACCAGGGCAAGGAGCTGGACTGCAGGGGAGGC 1908

Qy   2086 catgtagttttagatgatgatacggatgtgggctatgtagaagatggaacgccatgtggc 2145
          || ||      | |  || |   | ||  || ||||||| || ||||| ||  | || ||
Db   1909 CACGTGCAGCTGGCGGACGGCTCTGACCTGAGCTATGTGGAGGATGGCACAGCCTGCGGG 1968

Qy   2146 ccgtctatgatgtgtttagatcggaagtgcctacaaattcaagccctaaatatgagcagc 2205
```

```
Db   1969 CCTAACATGTTGTGCCTGGACCATCGCTGCCTGCCAGCTTCTGCCTTCAACTTCAGCACC 2028

Qy   2206 tgtccactcgattccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagcc 2265
Db   2029 TGCCCCGGCAGTGGGGAGCGCCGGATTTGCTCCCACCACGGGGTCTGCAGCAATGAAGGG 2088

Qy   2266 acctgcatttgtgatttcacctgggcagggacagattgcagtatccgggatccagttagg 2325
Db   2089 AAGTGCATCTGTCAGCCAGACTGGACAGGCAAAGACTGCAGTATCCATAACCCCCTGCCC 2148

Qy   2326 aaccttcacccccccaaggat---gaaggacccaagggtcctagtgccaccaatctcata 2382
Db   2149 ACGTCCCCACCCACGGGGGGAGACGGAGAGATATAAAGGTCCCAGCGGCACCAACATCATC 2208

Qy   2383 ataggctccatcgctggtgccatcctggtagcagctattgtccttgggggcacaggctgg 2442
Db   2209 ATTGGCTCCATCGCTGGGGCTGTCCTGGTTGCAGCCATCGTCCTGGGCGGCACGGGCTGG 2268

Qy   2443 ggatttaaaaatgtcaagaagagaaggttcg 2473
Db   2269 GGATTTAAAAACATTCGCCGAGGAAGGTCCG 2299
```

```
RESULT   4
US-08-243-542-7
; Sequence 7, Application US/08243542
; Patent No. 5552526
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  :  EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/243,542
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
       APPLICATION NUMBER:  JP 5-136602
       FILING DATE:  14 MAY 1993
       APPLICATION NUMBER:  JP 5-257455
       FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
;      TELEFAX:  (616) 381-5465
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2913 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY:  human fetal brain cDNA library
;    FEATURE:
;      NAME/KEY:  5' UTR
;      LOCATION:  1..27
;    FEATURE:
;      NAME/KEY:  3' UTR
;      LOCATION:  2038..2913
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  28..2037
US-08-243-542-7
```

Query Match 21.2%; Score 530; DB 1; Length 2913;
Best Local Similarity 57.8%; Pred. No. 9.7e-139;
Matches 1114; Conservative 0; Mismatches 775; Indels 39; Gaps 8;

```
Qy    573 gggtggagagcactgttactaccatggaagcatcagaggcgtcaaagactccaaggtggc 632
Db    114 GGCTGGAGACCACTGCTACTACCAGGGGAAGCTCCGGGGGAACCCGCACTCCTTCGCCGC 173

Qy    633 tctgtcaacctgcaatggacttcatggcatgtttgaagatgataccttcgtgtatatgat 692
Db    174 CCTCTCCACCTGCCAGGGGCTGCATGGGGTCTTCTCTGATGGGAACTTGACTTACATCGT 233

Qy    693 agagccactagagctggttcatgatgagaaaagcac------aggtcgaccacatataat 746
Db    234 GGAGCCCCAAGAGGTGGCTGGACCTTGGGGAGCCCCTCAGGGACCCCTTCCCCACCTCAT 293

Qy    747 ccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcactatggaaagagg 806
Db    294 TTACCGGACCCCTCTCCTCCCAGATCCCCTCGGATGCAGGGAACCAGGCTGCCTGTTTGC 353

Qy    807 tgaccagtggccctttctctctgaattacagtggttgaaaagaaggaagagagcagtgaa 866
Db    354 TGTGCCTGCCCAGTCGGCTCCTCCAAACCGGCCGAGGCTGAGAAGGAAAAGGCAGGTCCG 413

Qy    867 tccatcacgtggtatatttga---agaaatgaaatatttggaacttatgattgttaatga 923
Db    414 CCGGGGCCACCCTACAGTGCACAGTGAAACCAAGTATGTGGAGCTAATTGTGATCAACGA 473

Qy    924 tcacaaaacgtataagaagcatcgctcttctcatgcacataccaacaactttgcaaagtc 983
Db    474 CCACCAGCTGTTCGAGCAGATGCGACAGTCGGTGGTCCTCACCAGCAACTTTGCCAAGTC 533

Qy    984 cgtggtcaaccttgtggattctatttacaaggagcagctcaacaccagggttgtcctggt 1043
Db    534 CGTGGTGAACCTGGCCGATGTGATATACAAGGAGCAGCTCAACACTCGCATCGTCCTGGT 593

Qy   1044 ggctgtagagacctggactgagaaggatcagattgacatcaccaccaaccctgtgcagat 1103
Db    594 TGCCATGGAAACATGGGCAGATGGGGACAAGATCCAGGTGCAGGATGACCTCCTGGAGAC 653

Qy   1104 gctccatgagttctcaaaatac---cggcagcgcattaagcagcatgctgatgctgtgca 1160
Db    654 CCTGGCCCGGCTCATGGTCTACCGACGGGAGGGTCTGCCTGAGCCCAGTAATGCCACCCA 713

Qy   1161 cctcatctcgcgggtgacatttcactataagagaagcagtctgagttactttggaggtgt 1220
Db    714 CCTCTTCTCGGGCAGGACCTTCCAGAGCACGAGCAGCGGGGCAGCCTACGTGGGGGGCAT 773

Qy   1221 ctgttctcgcacaagaggagttggtgtgaatgagtatggtcttccaatggcagtggcaca 1280
Db    774 ATGCTCCCTGTCCCATGGCGGGGGTGTGAACGAGTACGGCAACATGGGGGCGATGGCCGT 833

Qy   1281 agtattatcgcagagcctggctcaaaaccttggaatccaatgggaaccttctagcagaaa 1340
Db    834 GACCCTTGCCCAGACGCTGGGACAGAACCTGGGCATGATGTGGAACAAACACCGGAGCTC 893
```

```
Qy   1341 gcca---aaatgtgactgcacagaatcctggggtggctgcatcatggaggaaacaggggt 1397
          | ||  .  | ||  | ||  ||||   ||||   ||||||||||||||||| || ||| |
Db    894 GGCAGGGGACTGCAAGTGTCCAGACATCTGGCTGGGCTGCATCATGGAGGACACTGGGTT 953

Qy   1398 gtcccattctcgaaaattttcaaagtgcagcattttggagtatagagacttttacagag 1457
          | ||   | || || || ||    ||||||||    ||||| |   | ||| | |||
Db    954 CTACCTGCCCCGCAAGTTCTCTCGCTGCAGCATCGACGAGTACAACCAGTTTCTGCAGGA 1013

Qy   1458 aggaggtggagcctgccttttcaacaggccaacaaagctatttgagcccacggaatgtgg 1517
           || |||||   ||||||  |||||||  |||    |||||  | || ||| | || || ||
Db   1014 GGGTGGTGGCAGCTGCCTCTTCAACAAGCCCCTCAAGCTCCTGGACCCCCCAGAGTGCGG 1073

Qy   1518 aaatggatacgtggaagctggggaggagtgtgattgtggttttcatgtggaatgc----- 1572
           || || | |||||| || ||||||||||| || || || |       ||| |||
Db   1074 GAACGGCTTCGTGGAGGCAGGGGAGGAGTGCGACTGCGGCTCGGTGCAGGAGTGCAGCCG 1133

Qy   1573 ----tatggattatgctgtaagaaatgttccctctccaacggggctcactgcagcgacgg 1628
               |||   |||| |||||||  ||||  |  |||  ||    ||||||||||
Db   1134 CGCAGGTGGCAACTGCTGCAAGAAATGCACCCTGACTCACGACGCCATGTGCAGCGACGG 1193

Qy   1629 gccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgccgggatgc 1688
           || |||||||  |          ||    |   | |||||  ||        |||||  || ||
Db   1194 GCTCTGCTGTCGCCGC------TGCAAGTACGAACCACGGGGTGTGTCCTGCCGAGAGGC 1247

Qy   1689 tgtgaacgagtgtgatattactgaatattgtactggagactctggtcagtgcccaccaaa 1748
           ||||||||||| || ||  | ||    || || || ||||||  | |||||||| || ||
Db   1248 CGTGAACGAGTGCGACATCGCGGAGACCTGCACCGGGGACTCTAGCCAGTGCCCGCCTAA 1307

Qy   1749 tcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctacaatggcga 1808
           || || ||||  ||||| ||    ||  | ||  | ||||||||||||||   ||
Db   1308 CCTGCACAAGCTGGACGGTTACTACTGTGACCATGAGCAGGGCCGCTGCTACGGAGGTCG 1367

Qy   1809 gtgcaagaccagagacaaccagtgtcagtacatctggggaacaaagcctgcagggtctga 1868
           ||||| ||| | |||   ||||| |||    | |||||      |||||      ||||
Db   1368 CTGCAAAACCCGGGACCGGCAGTGCCAGGTTCTTTGGGGCCATGCGGCTG------CTGA 1421

Qy   1869 caagttctgctatgaaaagctgaatacagaaggcactgagaagggaaactgcgggaagga 1928
              |||||||| || |||||||||   || || || |||   || | ||| |||    |
Db   1422 TCGCTTCTGCTACGAGAAGCTGAATGTGGAGGGGACGGAGCGTGGGAGCTGTGGGCGCAA 1481

Qy   1929 tggagaccggtggattcagtgcagcaaacatgatgtgttctgtggattcttactctgtac 1988
           |||  | | ||| | |||||||| || || || ||| | ||||| ||| | ||||||
Db   1482 GGGATCCGGCTGGGTCCAGTGCAGTAAGCAGGACGTGCTGTGTGGCTTCCTCCTCTGTGT 1541

     1989 caatcttactcgagctccacgtattggtcaacttcagggtgagatcattccaacttcctt 2048
          |||  |  || |||||||  ||  | ||  | ||   ||| || |||| |       ||||
     1542 CAACATCTCTGGAGCTCCTCGGCTAGGGGACCTGGTGGGAGACATCAGTAGTGTCACCTT 1601

Qy   2049 ctaccatcaaggccgggtgattgactgcagtggtgcccatgtagtttatagatgatgatac 2108
          |||||| || |||  || | | |||||||| || | ||| ||      | |  || |   |
Db   1602 CTACCACCAGGGCAAGGAGCTGGACTGCAGGGGAGGCCACGTGCAGCCTGGCGGACGGCTC 1661

Qy   2109 ggatgtgggctatgtagaagatggaacgccatgtggcccgtctatgatgtgtttagatcg 2168
           ||  || |||||| || ||||| ||  | || || ||    ||| ||||  | || |
Db   1662 TGACCTGAGCTATGTGGAGGATGGCACAGCCTGCGGGCCTAACATGTTGTGCCTGGACCA 1721

Qy   2169 gaagtgcctacaaattcaagccctaaatatgagcagctgtccactccattccaagggtaa 2228
          |||||  | |  |  ||| | ||  | |||| ||| ||   |  |    || | |
Db   1722 TCGCTGCCTGCCAGCTTCTGCCTTCAACTTCAGCACCTGCCCCGGCAGTGGGGAGCGCCG 1781

Qy   2229 agtctgttcgggccatggggtgtgtagtaatgaagccacctgcatttgtgatttcacctg 2288
            | || ||   ||| ||||| || || |||||||  |  ||||| ||| |      |||
Db   1782 GATTTGCTCCCACCACGGGGTCTGCAGCAATGAAGGGAAGTGCATCTGTCAGCCAGACTG 1841

Qy   2289 ggcagggacagattgcagtatccgggatccagt---taggaaccttcaccccccaagga 2345
          | |||| | ||| ||||||||||   | ||  |     | |  ||  | | |     ||
Db   1842 GACAGGCAAAGACTGCAGTATCCATAACCCCCTGCCCACGTCCCCACCCACGGGGGAGAC 1901

Qy   2346 tgaaggacccaagggtcctagtgccaccaatctcataataggctccatcgctggtgccat 2405
           ||  ||   || ||||| || | ||||||  |||| || ||||||||||||| ||  |
Db   1902 GGAGAGATATAAAGGTCCCAGCGGCACCAACATCATCATTGGCTCCATCGCTGGGGCTGT 1961
```

```
Qy   2406 cctggtagcagctattgtccttgggggcacaggctggggatttaaaaatgtcaagaagag 2465
          ||||||  |||||    |||||  || |||||  ||||||||||||||||  |       |
Db   1962 CCTGGTTGCAGCCTACGTCCTGGGCGGCACGGGCTGGGGATTTAAAAACATTCGCCGAGG 2021

Qy   2466 aaggttcg 2473
          |||||  ||
Db   2022 AAGGTCCG 2029
```

```
RESULT   5
US-08-477-407-7
; Sequence 7, Application US/08477407
; Patent No. 5631351
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  :  EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,407
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/243,542
;      FILING DATE:  13-MAY-1994
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
;      TELEFAX:  (616) 381-5465
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2913 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY:  human fetal brain cDNA library
;    FEATURE:
;      NAME/KEY:  5' UTR
;      LOCATION:  1..27
;    FEATURE:
;      NAME/KEY:  3' UTR
;      LOCATION:  2038..2913
;    FEATURE:
```

```
;       NAME/KEY:  CDS
;       LOCATION:  28..2037
US-08-477-407-7


  Query Match             21.2%;  Score 530;  DB 1;  Length 2913;
  Best Local Similarity   57.8%;  Pred. No. 9.7e-139;
  Matches 1114;  Conservative    0;  Mismatches  775;  Indels   39;  Gaps    8;

Qy     573 gggtggagagcactgttactaccatggaagcatcagaggcgtcaaagactccaaggtggc 632
           || ||||||| ||||| |||||||| || |   || | ||   |    |||||   |  ||
Db     114 GGCTGGAGACCACTGCTACTACCAGGGGAAGCTCCGGGGGAACCCGCACTCCTTCGCCGC 173

Qy     633 tctgtcaacctgcaatggacttcatggcatgtttgaagatgataccttcgtgtatatgat 692
           || || |||||| | || || |||||  | ||    ||||  | |||    || ||  |
Db     174 CCTCTCCACCTGCCAGGGGCTGCATGGGGTCTTCTCTGATGGGAACTTGACTTACATCGT 233

Qy     693 agagccactagagctggttcatgatgagaaaagcac------aggtcgaccacatataat 746
           ||||| | |||| ||| |     |  |  |  | |       |   |  || ||  | ||
Db     234 GGAGCCCCAAGAGGTGGCTGGACCTTGGGGAGCCCCTCAGGGACCCCTTCCCCACCTCAT 293

       747 ccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcactatggaaagagg 806
             |   |||          |   || |       |   | | | |               |
       294 TTACCGGACCCCTCTCCTCCCAGATCCCCTCGGATGCAGGGAACCAGGCTGCCTGTTTGC 353

Qy     807 tgaccagtggccctttctctctgaattacagtggttgaaaagaaggaagagagcagtgaa 866
           ||  |    | |       ||  |   | |  | |   |||||||| ||     ||
Db     354 TGTGCCTGCCCAGTCGGCTCCTCCAAACCGGCCGAGGCTGAGAAGGAAAAGGCAGGTCCG 413

Qy     867 tccatcacgtggtatatttga---agaaatgaaatatttggaacttatgattgttaatga 923
           |     |    || | |  |    ||||  || ||| |||| || ||  |  | || ||
Db     414 CCGGGGCCACCCTACAGTGCACAGTGAAACCAAGTATGTGGAGCTAATTGTGATCAACGA 473

Qy     924 tcacaaaacgtataagaagcatcgctcttctcatgcacataccaacaactttgcaaagtc 983
           ||| |   ||   || ||   ||     |  | |||| |||||||| |||||
Db     474 CCACCAGCTGTTCGAGCAGATGCGACAGTCGGTGGTCCTCACCAGCAACTTTGCCAAGTC 533

Qy     984 cgtggtcaaccttgtggattctatttacaaggagcagctcaacaccagggttgtcctggt 1043
           |||||| ||||| |  |||    || ||||||||||||||||||||  |  | ||||||||
Db     534 CGTGGTGAACCTGGCCGATGTGATATACAAGGAGCAGCTCAACACTCGCATCGTCCTGGT 593

Qy    1044 ggctgtagagacctggactgagaaggatcagattgacatcaccaccaaccctgtgcagat 1103
           ||  | || || ||| | ||   |||  ||||  |  |        |||   || |||
Db     594 TGCCATGGAAACATGGGCAGATGGGGACAAGATCCAGGTGCAGGATGACCTCCTGGAGAC 653

Qy    1104 gctccatgagttctcaaaatac---cggcagcgcattaagcagcatgctgatgctgtgca 1160
           ||       | ||      |||   ||| || | |     |||    | ||||     ||
Db     654 CCTGGCCCGGCTCATGGTCTACCGACGGGAGGGTCTGCCTGAGCCCAGTAATGCCACCCA 713

Qy    1161 cctcatctcgcgggtgacatttcactataagagaagcagtctgagttactttggaggtgt 1220
           |||| ||||| |   ||| || ||    | ||| ||| |       ||| | || ||  |
Db     714 CCTCTTCTCGGGCAGGACCTTCCAGAGCACGAGCAGCGGGGCAGCCTACGTGGGGGGCAT 773

Qy    1221 ctgttctcgcacaagaggagttggtgtgaatgagtatggtcttccaatggcagtggcaca 1280
           || || |   |    || |  |||||||| ||||| ||       |||  ||||
Db     774 ATGCTCCCTGTCCCATGGCGGGGGTGTGAACGAGTACGGCAACATGGGGGCGATGGCCGT 833

Qy    1281 agtattatcgcagagcctggctcaaaaccttggaatccaatgggaaccttctagcagaaa 1340
             |  | ||||  |||| || ||||| || ||    ||| |          | ||
Db     834 GACCCTTGCCCAGACGCTGGGACAGAACCTGGGCATGATGTGGAACAAACACCGGAGCTC 893

Qy    1341 gcca---aaatgtgactgcacagaatcctggggtggctgcatcatggaggaaacaggggt 1397
           | ||    | ||  | ||  ||||   ||||  ||||||||||||||||| || ||| |
Db     894 GGCAGGGGACTGCAAGTGTCCAGACATCTGGCTGGGCTGCATCATGGAGGACACTGGGTT 953

Qy    1398 gtcccattctcgaaaattttcaaagtgcagcattttggagtatagagacttttttacagag 1457
           | || · | || || || ||    ||||||||    |||||  |   | ||| | |||
Db     954 CTACCTGCCCCGCAAGTTCTCTCGCTGCAGCATCGACGAGTACAACCAGTTTCTGCAGGA 1013

Qy    1458 aggaggtggagcctgccttttcaacaggccaacaaagctatttgagcccacggaatgtgg 1517
           || |||||   |||||| ||||||| |||    |||||  | || ||| | || || ||
Db    1014 GGGTGGTGGCAGCTGCCTCTTCAACAAGCCCCTCAAGCTCCTGGACCCCCCAGAGTGCGG 1073

Qy    1518 aaatggatacgtggaagctggggaggagtgtgattgtggttttcatgtggaatgc----- 1572
           || || | |||||| || ||||||||||| || || || |       ||| |||
Db    1074 GAACGGCTTCGTGGAGGCAGGGGAGGAGTGCGACTGCGGCTCGGTGCAGGAGTGCAGCCG 1133

Qy    1573 ----tatggattatgctgtaagaaatgttccctctccaacggggctcactgcagcgacgg 1628
               |||    ||||| ||||||||  ||||  |  |||  ||    ||||||||||
Db    1134 CGCAGGTGGCAACTGCTGCAAGAAATGCACCCTGACTCACGACGCCATGTGCAGCGACGG 1193

Qy    1629 gccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgccgggatgc 1688
           || |||||||  |          ||    |  ||||| ||        |||||  || ||
Db    1194 GCTCTGCTGTCGCCGC------TGCAAGTACGAACCACGGGGTGTGTCCTGCCGAGAGGC 1247

Qy    1689 tgtgaacgagtgtgatattactgaatattgtactggagactctggtcagtgcccaccaaa 1748
           ||||||||||| || ||  | ||     || || || |||||| | |||||||| || ||
Db    1248 CGTGAACGAGTGCGACATCGCGGAGACCTGCACCGGGGACTCTAGCCAGTGCCCGCCTAA 1307

Qy    1749 tcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctacaatggcga 1808
           || || ||||  |||||  ||    ||  | ||  | ||||||||||||||   ||
Db    1308 CCTGCACAAGCTGGACGGTTACTACTGTGACCATGAGCAGGGCCGCTGCTACGGAGGTCG 1367

Qy    1809 gtgcaagaccagagacaaccagtgtcagtacatctggggaacaaaggctgcagggtctga 1868
           |||||  ||| | |||   |||||  |||    | |||||      |||||       ||||
Db    1368 CTGCAAAACCCGGGACCGGCAGTGCCAGGTTCTTTGGGGCCATGCGGCTG------CTGA 1421

Qy    1869 caagttctgctatgaaaagctgaatacagaaggcactgagaagggaaactgcgggaagga 1928
           |||||||| || |||||||||    || || || |||    || | ||| |||    |
Db    1422 TCGCTTCTGCTACGAGAAGCTGAATGTGGAGGGGACGGAGCGTGGGAGCTGTGGGCGCAA 1481

Qy    1929 tggagaccggtggattcagtgcagcaaacatgatgtgttctgtggattcttactctgtac 1988
           |||  | | ||| | |||||||| || || || ||| | ||||| ||| | ||||||
Db    1482 GGGATCCGGCTGGGTCCAGTGCAGTAAGCAGGACGTGCTGTGTGGCTTCCTCCTCTGTGT 1541

Qy    1989 caatcttactcgagctccacgtattggtcaacttcagggtgagatcattccaacttcctt 2048
           |||  |  || ||||||| ||  | ||  | ||   ||| || |||| |       ||||
Db    1542 CAACATCTCTGGAGCTCCTCGGCTAGGGGACCTGGTGGGAGACATCAGTAGTGTCACCTT 1601

Qy    2049 ctaccatcaaggccgggtgattgactgcagtggtgcccatgtagttttagatgatgatac 2108
           |||||| || |||  || | | |||||||| || | ||| ||      | |  || |   |
Db    1602 CTACCACCAGGGCAAGGAGCTGGACTGCAGGGGAGGCCACGTGCAGCTGGCGGACGGCTC 1661

Qy    2109 ggatgtgggctatgtagaagatggaacgccatgtggcccgtctatgatgtgtttagatcg 2168
           ||  || ||||||| || ||||| ||  | || || ||    ||| ||||  | || |
Db    1662 TGACCTGAGCTATGTGGAGGATGGCACAGCCTGCGGGCCTAACATGTTGTGCCTGGACCA 1721

Qy    2169 gaagtgcctacaaattcaagccctaaatatgagcagctgtccactcgattccaagggtaa 2228
           |||||  | |  |   ||| | ||  | |||| ||| ||    |  |    || |
Db    1722 TCGCTGCCTGCCAGCTTCTGCCTTCAACTTCAGCACCTGCCCCGGCAGTGGGGAGCGCCG 1781

Qy    2229 agtctgttcgggccatggggtgtgtagtaatgaagccacctgcatttgtgatttcacctg 2288
             | || ||   ||| ||||| || || |||||||  |  |||||  ||| |      |||
Db    1782 GATTTGCTCCCACCACGGGGTCTGCAGCAATGAAGGGAAGTGCATCTGTCAGCCAGACTG 1841

Qy    2289 ggcagggacagattgcagtatccgggatccagt---taggaaccttcacccccccaagga 2345
           | |||| | ||| ||||||||||   | ||  |   | |  ||  | | |     ||
Db    1842 GACAGGCAAAGACTGCAGTATCCATAACCCCCTGCCCACGTCCCCACCCACGGGGGAGAC 1901

Qy    2346 tgaaggacccaagggtcctagtgccaccaatctcataataggctccatcgctggtgccat 2405
           ||  ||   || ||||| || | ||||||  |||| || |||||||||||||| ||  |
Db    1902 GGAGAGATATAAAGGTCCCAGCGGCACCAACATCATCATTGGCTCCATCGCTGGGGCTGT 1961

Qy    2406 cctggtagcagctattgtccttgggggcacaggctggggatttaaaaatgtcaagaagag 2465
           ||||||  |||||    ||||| || |||||  ||||||||||||||||  |       |
Db    1962 CCTGGTTGCAGCCTACGTCCTGGGCGGCACGGGCTGGGGATTTAAAAACATTCGCCGAGG 2021

Qy    2466 aaggttcg 2473
           |||||  ||
Db    2022 AAGGTCCG 2029


RESULT   6
US-08-484-355-7
; Sequence 7, Application US/08484355
; Patent No. 5705341
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  :  EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,355
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/243,542
;      FILING DATE:  13-MAY-1994
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
;      TELEFAX:  (616) 381-5465
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2913 base pairs
       TYPE:  nucleic acid
       STRANDEDNESS:  double
       TOPOLOGY:  linear
    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY:  human fetal brain cDNA library
;    FEATURE:
;      NAME/KEY:  5' UTR
;      LOCATION:  1..27
;    FEATURE:
;      NAME/KEY:  3' UTR
;      LOCATION:  2038..2913
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  28..2037
US-08-484-355-7


  Query Match             21.2%;  Score 530;  DB 1;  Length 2913;
  Best Local Similarity   57.8%;  Pred. No. 9.7e-139;
  Matches 1114;  Conservative    0;  Mismatches  775;  Indels   39;  Gaps    8;

Qy    573 gggtggagagcactgttactaccatggaagcatcagaggcgtcaaagactccaaggtggc 632
          || |||||| ||||| |||||||| || |   || | ||   |    |||||   |  ||
Db    114 GGCTGGAGACCACTGCTACTACCAGGGGAAGCTCCGGGGGAACCCGCACTCCTTCGCCGC 173

Qy    633 tctgtcaacctgcaatggacttcatggcatgtttgaagatgataccttcgtgtatatgat 692
          || || |||||| | || || |||||  | ||    ||||  | |||    || ||  |
Db    174 CCTCTCCACCTGCCAGGGGCTGCATGGGGTCTTCTCTGATGGGAACTTGACTTACATCGT 233

Qy    693 agagccactagagctggttcatgatgagaaaagcac------aggtcgaccacatataat 746
          |||||  | ||||  ||| |     |  |  |  | |      |   |  || ||  | ||
Db    234 GGAGCCCCAAGAGGTGGCTGGACCTTGGGGAGCCCCTCAGGGACCCCTTCCCCACCTCAT 293

Qy    747 ccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcactatggaaagagg 806
            |    |||          |   || |      |   | | |  |              |
Db    294 TTACCGGACCCCTCTCCTCCCAGATCCCCTCGGATGCAGGGAACCAGGCTGCCTGTTTGC 353

Qy    807 tgaccagtggccctttctctctgaattacagtggttgaaaagaaggaagagagcagtgaa 866
          ||  |     |  |      ||  |   | |  |  |   |||||||| ||    ||
Db    354 TGTGCCTGCCCAGTCGGCTCCTCCAAACCGGCCGAGGCTGAGAAGGAAAAGGCAGGTCCG 413

Qy    867 tccatcacgtggtatatttga---agaaatgaaatatttggaacttatgattgttaatga 923
          |      |    || | |  |    ||||  || ||| |||| || ||  |  | || ||
Db    414 CCGGGGCCACCCTACAGTGCACAGTGAAACCAAGTATGTGGAGCTAATTGTGATCAACGA 473

Qy    924 tcacaaaacgtataagaagcatcgctcttctcatgcacataccaacaactttgcaaagtc 983
          |||  |   ||   || ||   ||    ||    |  |  |||| |||||||| |||||
Db    474 CCACCAGCTGTTCGAGCAGATGCGACAGTCGGTGGTCCTCACCAGCAACTTTGCCAAGTC 533

Qy    984 cgtggtcaaccttgtggattctatttacaaggagcagctcaacaccagggttgtcctggt 1043
          ||||||  ||||| |  |||   || ||||||||||||||||||  |  | ||||||||
Db    534 CGTGGTGAACCTGGCCGATGTGATATACAAGGAGCAGCTCAACACTCGCATCGTCCTGGT 593

Qy   1044 ggctgtagagacctggactgagaaggatcagattgacatcaccaccaaccctgtgcagat 1103
          ||  | ||  ||  ||| | ||   |||  ||||  |  |         |||   || |||
Db    594 TGCCATGGAAACATGGGCAGATGGGGACAAGATCCAGGTGCAGGATGACCTCCTGGAGAC 653

Qy   1104 gctccatgagttctcaaaatac---cggcagcgcattaagcagcatgctgatgctgtgca 1160
          ||     | ||      |||   ||| || |  |     |||    | ||||    ||
Db    654 CCTGGCCCGGCTCATGGTCTACCGACGGGAGGGTCTGCCTGAGCCCAGTAATGCCACCCA 713

Qy   1161 cctcatctcgcgggtgacatttcactataagagaagcagtctgagttactttggaggtgt 1220
          ||||  |||||  |   |||  || ||    | |||  |||  |        |||  | ||  ||  |
Db    714 CCTCTTCTCGGGCAGGACCTTCCAGAGCACGAGCAGCGGGGCAGCCTACGTGGGGGGCAT 773

Qy   1221 ctgttctcgcacaagaggagttggtgtgaatgagtatggtcttccaatggcagtggcaca 1280
          ||  ||  |    |    ||  |  |||||||| |||||  ||          |||  ||||
Db    774 ATGCTCCCTGTCCCATGGCGGGGGTGTGAACGAGTACGGCAACATGGGGGCGATGGCCGT 833

Qy   1281 agtattatcgcagagcctggctcaaaaccttggaatccaatgggaaccttctagcagaaa 1340
               |  | ||||  ||||  || |||||  || ||    |||  |         | ||
Db    834 GACCCTTGCCCAGACGCTGGGACAGAACCTGGGCATGATGTGGAACAAACACCGGAGCTC 893

Qy   1341 gcca---aaatgtgactgcacagaatcctggggtggctgcatcatggaggaaacaggggt 1397
          | ||    | ||  | ||  ||||   ||||  ||||||||||||||||||  || ||| |
Db    894 GGCAGGGGACTGCAAGTGTCCAGACATCTGGCTGGGCTGCATCATGGAGGACACTGGGTT 953

Qy   1398 gtcccattctcgaaaattttcaaagtgcagcattttggagtatagagacttttttacagag 1457
          | ||   | || || || ||    ||||||||    |||||  |   | ||| | |||
Db    954 CTACCTGCCCCGCAAGTTCTCTCGCTGCAGCATCGACGAGTACAACCAGTTTCTGCAGGA 1013

Qy   1458 aggaggtggagcctgccttttcaacaggccaacaaagctatttgagcccacggaatgtgg 1517
          || |||||   |||||| |||||||  |||    |||||  | || ||| | || || ||
Db   1014 GGGTGGTGGCAGCTGCCTCTTCAACAAGCCCCTCAAGCTCCTGGACCCCCCAGAGTGCGG 1073

Qy   1518 aaatggatacgtggaagctggggaggagtgtgattgtggttttcatgtggaatgc----- 1572
          || || | |||||| || |||||||||| || || || |       ||| |||
Db   1074 GAACGGCTTCGTGGAGGCAGGGGAGGAGTGCGACTGCGGCTCGGTGCAGGAGTGCAGCCG 1133

Qy   1573 ----tatggattatgctgtaagaaatgttccctctccaacggggctcactgcagcgacgg 1628
              |||    ||||| ||||||||  ||||  |  |||  ||    ||||||||||
Db   1134 CGCAGGTGGCAACTGCTGCAAGAAATGCACCCTGACTCACGACGCCATGTGCAGCGACGG 1193

Qy   1629 gccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgccgggatgc 1688
          || |||||||  |         ||    |   | |||||  ||       |||||  || ||
Db   1194 GCTCTGCTGTCGCCGC------TGCAAGTACGAACCACGGGGTGTGTCCTGCCGAGAGGC 1247

Qy   1689 tgtgaacgagtgtgatattactgaatattgtactggagactctggtcagtgcccaccaaa 1748
```

```
Db    1248 CGTGAACGAGTGCGACATCGCGGAGACCTGCACCGGGGACTCTAGCCAGTGCCCGCCTAA 1307

Qy    1749 tcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctacaatggcga 1808
Db    1308 CCTGCACAAGCTGGACGGTTACTACTGTGACCATGAGCAGGGCCGCTGCTACGGAGGTCG 1367

Qy    1809 gtgcaagaccagagacaaccagtgtcagtacatctggggaacaaaggctgcagggtctga 1868
Db    1368 CTGCAAAACCCGGGACCGGCAGTGCCAGGTTCTTTGGGGCCATGCGGCTG------CTGA 1421

Qy    1869 caagttctgctatgaaaagctgaatacagaaggcactgagaagggaaactgcgggaagga 1928
Db    1422 TCGCTTCTGCTACGAGAAGCTGAATGTGGAGGGGACGGAGCGTGGGAGCTGTGGGCGCAA 1481

Qy    1929 tggagaccggtggattcagtgcagcaaacatgatgtgttctgtggattcttactctgtac 1988
Db    1482 GGGATCCGGCTGGGTCCAGTGCAGTAAGCAGGACGTGCTGTGTGGCTTCCTCCTCTGTGT 1541

Qy    1989 caatcttactcgagctccacgtattggtcaacttcagggtgagatcattccaacttcctt 2048
      1542 CAACATCTCTGGAGCTCCTCGGCTAGGGGACCTGGTGGGAGACATCAGTAGTGTCACCTT 1601

      2049 ctaccatcaaggccgggtgattgactgcagtggtgcccatgtagttttagatgatgatac 2108
Db    1602 CTACCACCAGGGCAAGGAGCTGGACTGCAGGGGAGGCCACGTGCAGCTGGCGGACGGCTC 1661

Qy    2109 ggatgtgggctatgtagaagatggaacgccatgtggcccgtctatgatgtgtttagatcg 2168
Db    1662 TGACCTGAGCTATGTGGAGGATGGCACAGCCTGCGGGCCTAACATGTTGTGCCTGGACCA 1721

Qy    2169 gaagtgcctacaaattcaagccctaaatatgagcagctgtccactcgattccaagggtaa 2228
Db    1722 TCGCTGCCTGCCAGCTTCTGCCTTCAACTTCAGCACCTGCCCCGGCAGTGGGGAGCGCCG 1781

Qy    2229 agtctgttcgggccatggggtgtgtagtaatgaagccacctgcatttgtgatttcacctg 2288
Db    1782 GATTTGCTCCCACCACGGGGTCTGCAGCAATGAAGGGAAGTGCATCTGTCAGCCAGACTG 1841

Qy    2289 ggcagggacagattgcagtatccgggatccagt---taggaaccttcacccccccaagga 2345
Db    1842 GACAGGCAAAGACTGCAGTATCCATAACCCCCTGCCCACGTCCCCACCCACGGGGGAGAC 1901

Qy    2346 tgaaggacccaagggtcctagtgccaccaatctcataataggctccatcgctggtgccat 2405
Db    1902 GGAGAGATATAAAGGTCCCAGCGGCACCAACATCATCATTGGCTCCATCGCTGGGGCTGT 1961

Qy    2406 cctggtagcagctattgtccttgggggcacaggctggggatttaaaaatgtcaagaagag 2465
Db    1962 CCTGGTTGCAGCCTACGTCCTGGGCGGCACGGGCTGGGGATTTAAAAACATTCGCCGAGG 2021

      2466 aaggttcg 2473
      2022 AAGGTCCG 2029
```

```
RESULT   7
US-08-243-542-6
; Sequence 6, Application US/08243542
; Patent No. 5552526
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  :  EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/243,542
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
;      TELEFAX:  (616) 381-5465
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2923 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY:  human fetal brain cDNA library
;    FEATURE:
;      NAME/KEY:  5' UTR
;      LOCATION:  1..27
;    FEATURE:
;      NAME/KEY:  3' UTR
;      LOCATION:  1600..2923
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  28..1599
US-08-243-542-6
```

Query Match 20.5%; Score 513.2; DB 1; Length 2923;
Best Local Similarity 57.6%; Pred. No. 5e-134;
Matches 1116; Conservative 0; Mismatches 773; Indels 49; Gaps 9;

```
Qy     573 gggtggagagcactgttactaccatggaagcatcagaggcgtcaaagactccaaggtggc 632
Db     114 GGCTGGAGACCACTGCTACTACCAGGGGAAGCTCCGGGGGAACCCGCACTCCTTCGCCGC 173

Qy     633 tctgtcaacctgcaatggacttcatggcatgtttgaagatgataccttcgtgtatatgat 692
Db     174 CCTCTCCACCTGCCAGGGGCTGCATGGGGTCTTCTCTGATGGGAACTTGACTTACATCGT 233

Qy     693 agagccactagagctggttcatgatgagaaaagcac------aggtcgaccacatataat 746
Db     234 GGAGCCCCAAGAGGTGGCTGGACCTTGGGGAGCCCCTCAGGGACCCCTTCCCCACCTCAT 293

Qy     747 ccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcactatggaaagagg 806
Db     294 TTACCGGACCCCTCTCCTCCCAGATCCCCTCGGATGCAGGGAACCAGGCTGCCTGTTTGC 353

Qy     807 tgaccagtggccctttctctctgaattacagtggttgaaaagaaggaagagagcagtgaa 866
Db     354 TGTGCCTGCCCAGTCGGCTCCTCCAAACCGGCCGAGGCTGAGAAGGAAAAGGCAGGTCCG 413
```

```
Qy    867 tccatcacgtggtatatttga---agaaatgaaatatttggaacttatgattgttaatga 923
Db    414 CCGGGGCCACCCTACAGTGCACAGTGAAACCAAGTATGTGGAGCTAATTGTGATCAACGA 473

Qy    924 tcacaaaacgtataagaagcatcgctcttctcatgcacataccaacaactttgcaaagtc 983
Db    474 CCACCAGCTGTTCGAGCAGATGCGACAGTCGGTGGTCCTCACCAGCAACTTTGCCAAGTC 533

Qy    984 cgtggtcaaccttgtggattctatttacaaggagcagctcaacaccagggttgtcctggt 1043
Db    534 CGTGGTGAACCTGGCCGATGTGATATACAAGGAGCAGCTCAACACTCGCATCGTCCTGGT 593

Qy   1044 ggctgtagagacctggactgagaaggatcagattgacatcaccaccaaccctgtgcagat 1103
Db    594 TGCCATGGAAACATGGGCAGATGGGGACAAGATCCAGGTGCAGGATCACCTCCTGGAGAC 653

Qy   1104 gctccatgagttctcaaaatac---cggcagcgcattaagcagcatgctgatgctgtgca 1160
Db    654 CCTGGCCCGGCTCATGGTCTACCGACGGGAGGGTCTGCCTGAGCCCAGTAATGCCACCCA 713

Qy   1161 cctcatctcgcgggtgacatttcactataagagaagcagtctgagttactttggaggtgt 1220
Db    714 CCTCTTCTCGGGCAGGACCTTCCAGAGCACGAGCAGCGGGGCAGCCTACCTGGGGGGCAT 773

Qy   1221 ctgttctcgcacaagaggagttggtgtgaatgagtatggtcttccaatggcagtggcaca 1280
Db    774 ATGCTCCCTGTCCCATGGCGGGGGTGTGAACGAGTACGGCAACATGCGGGCGATGGCCGT 833

Qy   1281 agtattatcgcagagcctggctcaaaaccttggaatccaatgggaacccttctagcagaaa 1340
Db    834 GACCCTTGCCCAGACGCTGGGACAGAACCTGGGCATGATGTGGAACAAACACCGGAGCTC 893

Qy   1341 gcca---aaatgtgactgcacagaatcctggggtggctgcatcatgcaggaaacaggggt 1397
Db    894 GGCAGGGGACTGCAAGTGTCCAGACATCTGGCTGGGCTGCATCATGCAGGACACTGGGTT 953

Qy   1398 gtcccattctcgaaaattttcaaagtgcagcattttggagtatagacactttttacagag 1457
Db    954 CTACCTGCCCCGCAAGTTCTCTCGCTGCAGCATCGACGAGTACAACCAGTTTCTGCAGGA 1013

Qy   1458 aggaggtggagcctgccttttcaacaggccaacaaagctatttgagcccacggaatgtgg 1517
Db   1014 GGGTGGTGGCAGCTGCCTCTTCAACAAGCCCCTCAAGCTCCTGGACCCCCAGAGTGCGG 1073

     1518 aaatggatacgtggaagctggggaggagtgtgattgtggttttcatgtggaatgc----- 1572
     1074 GAACGGCTTCGTGGAGGCAGGGGAGGAGTGCGACTGCGGCTCGGTGCAGGAGTGCAGCCG 1133

Qy   1573 ----tatggattatgctgtaagaaatgttccctctccaacggggctcactgcagcgacgg 1628
Db   1134 CGCAGGTGGCAACTGCTGCAAGAAATGCACCCTGACTCACGACGCCATGTGCAGCGACGG 1193

Qy   1629 gccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgccgggatgc 1688
Db   1194 GCTCTGCTGTCGCCGC------TGCAAGTACGAACCACGGGGTGTGTCCTGCCGAGAGGC 1247

Qy   1689 tgtgaacgagtgtgatattactgaatattgtactggagactctggtcagtgcccaccaaa 1748
Db   1248 CGTGAACGAGTGCGACATCGCGGAGACCTGCACCGGGGACTCTAGCCAGTGCCCGCCTAA 1307

Qy   1749 tcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctacaatggcga 1808
Db   1308 CCTGCACAAGCTGGACGGTTACTACTGTGACCATGAGCAGGGCCGCTGCTACGGAGGTCG 1367

Qy   1809 gtgcaagaccagagacaaccagtgtcagtacatctggggaacaaagcctgcagggtctga 1868
Db   1368 CTGCAAAACCCGGGACCGGCAGTGCCAGGTTCTTTGGGGCCATGCGGCTG------CTGA 1421

Qy   1869 caagttctgctatgaaaagctgaatacagaaggcactgagaagggaactgcgggaagga 1928
Db   1422 TCGCTTCTGCTACGAGAAGCTGAATGTGGAGGGGACGGAGCGTGGGAGCTGTGGGCGCAA 1481

Qy   1929 tggagaccggtggattcagtgcagcaaaca----------tgatgtgttctgtggattct 1978
Db   1482 GGGATCCGGCTGGGTCCAGTGCAGTAAGCAGCCCCAACAGGGACGTGCTGTGTGGCTTCC 1541

Qy   1979 tactctgtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatcattc 2038
Db   1542 TCCTCTGTGTCAACATCTCTGGAGCTCCTCGGCTAGGGGACCTGGTGGGAGACATCAGTA 1601

Qy   2039 caacttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagttttag 2098
Db   1602 GTGTCACCTTCTACCACCAGGGCAAGGAGCTGGACTGCAGGGGAGGCCACGTGCAGCTGG 1661

Qy   2099 atgatgatacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatgatgt 2158
Db   1662 CGGACGGCTCTGACCTGAGCTATGTGGAGGATGGCACAGCCTGCGGGCCTAACATGTTGT 1721

Qy   2159 gtttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactcgatt 2218
Db   1722 GCCTGGACCATCGCTGCCTGCCAGCTTCTGCCTTCAACTTCAGCACCTGCCCCGGCAGTG 1781

Qy   2219 ccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatttgtg 2278
Db   1782 GGGAGCGCCGGATTTGCTCCCACCACGGGGTCTGCAGCAATGAAGGGAAGTGCATCTGTC 1841

Qy   2279 atttcacctgggcagggacagattgcagtatccgggatccagt---taggaaccttcacc 2335
Db   1842 AGCCAGACTGGACAGGCAAAGACTGCAGTATCCATAACCCCCTGCCCACGTCCCCACCCA 1901

Qy   2336 cccccaaggatgaaggacccaagggtcctagtgccaccaatctcataatággctccatcg 2395
Db   1902 CGGGGGAGACGGAGAGATATAAAGGTCCCAGCGGCACCAACATCATCATTGGCTCCATCG 1961

Qy   2396 ctggtgccatcctggtagcagctattgtccttgggggcacaggctggggatttaaaaatg 2455
Db   1962 CTGGGGCTGTCCTGGTTGCAGCCATCGTCCTGGGCGGCACGGGCTGGGGATTTAAAAACA 2021

Qy   2456 tcaagaagagaaggttcg 2473
Db   2022 TTCGCCGAGGAAGGTCCG 2039
```

```
RESULT   8
US-08-477-407-6
; Sequence 6, Application US/08477407
; Patent No. 5631351
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  :  EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,407
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/243,542
;      FILING DATE:  13-MAY-1994
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
```

```
;       FILING DATE:  22 SEPTEMBER 1993
;       APPLICATION NUMBER:  JP 6-49904
;       FILING DATE:  23 FEBRUARY 1994
;       APPLICATION NUMBER:  JP 6-73328
;       FILING DATE:  12 APRIL 1994
;       APPLICATION NUMBER:  JP 6-84470
;       FILING DATE:  22 APRIL 1994
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Terryence F. Chapman
;       REGISTRATION NUMBER:  32 549
;       REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (616) 381-1156
;       TELEFAX:  (616) 381-5465
;   INFORMATION FOR SEQ ID NO:  6:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2923 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
      MOLECULE TYPE:  cDNA to mRNA
      ORIGINAL SOURCE:
        ORGANISM:  Homo sapiens
      IMMEDIATE SOURCE:
,       LIBRARY:  human fetal brain cDNA library
;     FEATURE:
;       NAME/KEY:  5' UTR
;       LOCATION:  1..27
;     FEATURE:
;       NAME/KEY:  3' UTR
;       LOCATION:  1600..2923
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  28..1599
US-08-477-407-6

  Query Match             20.5%;  Score 513.2;  DB 1;  Length 2923;
  Best Local Similarity   57.6%;  Pred. No. 5e-134;
  Matches 1116;  Conservative    0;  Mismatches  773;  Indels   49;  Gaps    9;

Qy    573 gggtggagagcactgttactaccatggaagcatcagaggcgtcaaagactccaaggtggc 632
          || |||||| ||||| |||||||| || |   || | ||  |   |||||  |  ||
Db    114 GGCTGGAGACCACTGCTACTACCAGGGGAAGCTCCGGGGGAACCCGCACTCCTTCGCCGC 173

Qy    633 tctgtcaacctgcaatggacttcatggcatgtttgaagatgataccttcgtgtatatgat 692
          || || |||||| | || || |||||  | ||    ||||  | |||    || ||  |
Db    174 CCTCTCCACCTGCCAGGGGCTGCATGGGGTCTTCTCTGATGGGAACTTGACTTACATCGT 233

Qy    693 agagccactagagctggttcatgatgagaaaagcac------aggtcgaccacatataat 746
          |||||  | ||||  ||| |     |  |  | | |       |   |  || ||  | ||
Db    234 GGAGCCCCAAGAGGTGGCTGGACCTTGGGGAGCCCCTCAGGGACCCCTTCCCCACCTCAT 293

Qy    747 ccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcactatggaaagagg 806
            |    |||          |   || |       |   | | | |          |
Db    294 TTACCGGACCCCTCTCCTCCCAGATCCCCTCGGATGCAGGGAACCAGGCTGCCTGTTTGC 353

Qy    807 tgaccagtggccctttctctctgaattacagtggttgaaaagaaggaagagagcagtgaa 866
          ||  |     |  |       || |   | |  |  |   |||||||| ||    ||
Db    354 TGTGCCTGCCCAGTCGGCTCCTCCAAACCGGCCGAGGCTGAGAAGGAAAAGGCAGGTCCG 413

Qy    867 tccatcacgtggtatatttga---agaaatgaaatatttggaacttatgattgttaatga 923
          |      |    || | |  |    ||||  || ||| |||| || ||   |  | || ||
Db    414 CCGGGGCCACCCTACAGTGCACAGTGAAACCAAGTATGTGGAGCTAATTGTGATCAACGA 473

Qy    924 tcacaaaacgtataagaagcatcgctcttctcatgcacataccaacaactttgcaaagtc 983
          ||| |   ||   || ||   ||    ||    |  |  |||| |||||||| |||||
Db    474 CCACCAGCTGTTCGAGCAGATGCGACAGTCGGTGGTCCTCACCAGCAACTTTGCCAAGTC 533

Qy    984 cgtggtcaaccttgtggattctatttacaaggagcagctcaacaccagggttgtcctggt 1043
          ||||||  |||||  |  |||   || ||||||||||||||||||  |  | ||||||||
Db    534 CGTGGTGAACCTGGCCGATGTGATATACAAGGAGCAGCTCAACACTCGCATCGTCCTGGT 593

Qy   1044 ggctgtagagacctggactgagaaggatcagattgacatcaccaccaaccctgtgcagat 1103
          ||  | || || ||| | ||   |||  ||||  |  |         |||   || |||
Db    594 TGCCATGGAAACATGGGCAGATGGGGACAAGATCCAGGTGCAGGATGACCTCCTGGAGAC 653

Qy   1104 gctccatgagttctcaaaatac---cggcagcgcattaagcagcatgctgatgctgtgca 1160
          ||       | ||       |||   ||| || |  |      |||     | ||||    ||
Db    654 CCTGGCCCGGCTCATGGTCTACCGACGGGAGGGTCTGCCTGAGCCCAGTAATGCCACCCA 713

Qy   1161 cctcatctcgcgggtgacatttcactataagagaagcagtctgagttactttggaggtgt 1220
          |||| ||||| |   ||| || ||    | ||| ||| |         ||| | || ||  |
Db    714 CCTCTTCTCGGGCAGGACCTTCCAGAGCACGAGCAGCGGGGCAGCCTACGTGGGGGGCAT 773

Qy   1221 ctgttctcgcacaagaggagttggtgtgaatgagtatggtcttccaatggcagtggcaca 1280
          || || |   |    || |  |||||||| ||||| ||          |||  ||||
Db    774 ATGCTCCCTGTCCCATGGCGGGGGTGTGAACGAGTACGGCAACATGGGGGCGATGGCCGT 833

Qy   1281 agtattatcgcagagcctggctcaaaaccttggaatccaatgggaaccttctagcagaaa 1340
                |  | ||||  ||||  || |||||  || ||    |||  |         | ||
Db    834 GACCCTTGCCCAGACGCTGGGACAGAACCTGGGCATGATGTGGAACAAACACCGGAGCTC 893

Qy   1341 gcca---aaatgtgactgcacagaatcctggggtggctgcatcatggaggaaacaggggt 1397
          | ||    | ||  | ||  ||||   ||||   ||||||||||||||||| || ||| |
Db    894 GGCAGGGGACTGCAAGTGTCCAGACATCTGGCTGGGCTGCATCATGGAGGACACTGGGTT 953

Qy   1398 gtcccattctcgaaaattttcaaagtgcagcattttggagtatagagacttttttacagag 1457
          | ||   | || || || ||    |||||||    |||||  |   | ||| | |||
Db    954 CTACCTGCCCCGCAAGTTCTCTCGCTGCAGCATCGACGAGTACAACCAGTTTCTGCAGGA 1013

Qy   1458 aggaggtggagcctgccttttcaacaggccaacaaagctatttgagcccacggaatgtgg 1517
          || |||||   |||||| |||||| |||    |||||  | || ||| | || || ||
Db   1014 GGGTGGTGGCAGCTGCCTCTTCAACAAGCCCCTCAAGCTCCTGGACCCCCCAGAGTGCGG 1073

Qy   1518 aaatggatacgtggaagctggggaggagtgtgattgtggttttcatgtggaatgc----- 1572
          || || | |||||| || ||||||||||| || || || |     ||| |||
Db   1074 GAACGGCTTCGTGGAGGCAGGGGAGGAGTGCGACTGCGGCTCGGTGCAGGAGTGCAGCCG 1133

Qy   1573 ----tatggattatgctgtaagaaatgttccctctccaacggggctcactgcagcgacgg 1628
              |||    ||||| |||||||| |||| |  |||  ||    ||||||||||
Db   1134 CGCAGGTGGCAACTGCTGCAAGAAATGCACCCTGACTCACGACGCCATGTGCAGCGACGG 1193

Qy   1629 gccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgccgggatgc 1688
          || |||||||  |          ||    |  | ||||| ||       ||||| || ||
Db   1194 GCTCTGCTGTCGCCGC------TGCAAGTACGAACCACGGGGTGTGTCCTGCCGAGAGGC 1247

Qy   1689 tgtgaacgagtgtgatattactgaatattgtactggagactctggtcagtgcccaccaaa 1748
          ||||||||||| || ||  | ||    || || || |||||| | |||||||| || ||
Db   1248 CGTGAACGAGTGCGACATCGCGGAGACCTGCACCGGGGACTCTAGCCAGTGCCCGCCTAA 1307

Qy   1749 tcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctacaatggcga 1808
          || || |||| ||||| ||    || | || |  |||||||||||||||    ||
Db   1308 CCTGCACAAGCTGGACGGTTACTACTGTGACCATGAGCAGGGCCGCTGCTACGGAGGTCG 1367

Qy   1809 gtgcaagaccagagacaaccagtgtcagtacatctggggaacaaaggctgcagggtctga 1868
          ||||| ||| | |||   ||||| |||     | |||||       |||||       ||||
Db   1368 CTGCAAAACCCGGGACCGGCAGTGCCAGGTTCTTTGGGGCCATGCGGCTG------CTGA 1421

Qy   1869 caagttctgctatgaaaagctgaatacagaaggcactgagaagggaaactgcgggaagga 1928
            |||||||| || |||||||||   || || || |||   || | ||| |||    |
Db   1422 TCGCTTCTGCTACGAGAAGCTGAATGTGGAGGGGACGGAGCGTGGGAGCTGTGGGCGCAA 1481

Qy   1929 tggagaccggtggattcagtgcagcaaaca----------tgatgtgttctgtggattct 1978
          |||  | | ||| | |||||||| || ||           || ||| | ||||| |||
Db   1482 GGGATCCGGCTGGGTCCAGTGCAGTAAGCAGCCCCAACAGGGACGTGCTGTGTGGCTTCC 1541

Qy   1979 tactctgtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatcattc 2038
          | ||||||  |||  |  || |||||||  ||  | ||  | ||   ||| || |||| |
Db   1542 TCCTCTGTGTCAACATCTCTGGAGCTCCTCGGCTAGGGGACCTGGTGGGAGACATCAGTA 1601

Qy   2039 caacttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagttttag 2098
               |||||||||| || |||  || | | |||||||| || | ||| ||     | |
Db   1602 GTGTCACCTTCTACCACCAGGGCAAGGAGCTGGACTGCAGGGGAGGCCACGTGCAGCTGG 1661

Qy   2099 atgatgatacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatgatgt 2158
```

```
Db   1662 CGGACGGCTCTGACCTGAGCTATGTGGAGGATGGCACAGCCTGCGGGCCTAACATGTTGT 1721

Qy   2159 gtttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactcgatt 2218

Db   1722 GCCTGGACCATCGCTGCCTGCCAGCTTCTGCCTTCAACTTCAGCACCTGCCCCGGCAGTG 1781

Qy   2219 ccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatttgtg 2278

Db   1782 GGGAGCGCCGGATTTGCTCCCACCACGGGGTCTGCAGCAATGAAGGCAAGTGCATCTGTC 1841

Qy   2279 atttcacctgggcagggacagattgcagtatccgggatccagt---taggaaccttcacc 2335

Db   1842 AGCCAGACTGGACAGGCAAAGACTGCAGTATCCATAACCCCCTGCCCACGTCCCCACCCA 1901

Qy   2336 cccccaaggatgaaggacccaagggtcctagtgccaccaatctcataataggctccatcg 2395

Db   1902 CGGGGGAGACGGAGAGATATAAAGGTCCCAGCGGCACCAACATCATCATTGGCTCCATCG 1961

Qy   2396 ctggtgccatcctggtagcagctattgtccttgggggcacaggctgcggatttaaaaatg 2455

Db   1962 CTGGGGCTGTCCTGGTTGCAGCCATCGTCCTGGGCGGCACGGGCTGCGGATTTAAAAACA 2021

Qy   2456 tcaagaagagaaggttcg 2473

Db   2022 TTCGCCGAGGAAGGTCCG 2039


RESULT   9
US-08-484-355-6
; Sequence 6, Application US/08484355
; Patent No. 5705341
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  :  EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,355
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/243,542
;      FILING DATE:  13-MAY-1994
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
;      TELEFAX:  (616) 381-5465
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2923 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY:  human fetal brain cDNA library
;    FEATURE:
;      NAME/KEY:  5' UTR
;      LOCATION:  1..27
;    FEATURE:
;      NAME/KEY:  3' UTR
;      LOCATION:  1600..2923
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  28..1599
US-08-484-355-6


  Query Match             20.5%;  Score 513.2;  DB 1;  Length 2923;
  Best Local Similarity   57.6%;  Pred. No. 5e-134;
  Matches 1116;  Conservative    0;  Mismatches  773;  Indels   49;  Gaps    9;

Qy    573 gggtggagagcactgttactaccatggaagcatcagaggcgtcaaagactccaaggtggc 632

Db    114 GGCTGGAGACCACTGCTACTACCAGGGGAAGCTCCGGGGGAACCCGCACTCCTTCGCCGC 173

Qy    633 tctgtcaacctgcaatggacttcatggcatgtttgaagatgataccttcgtgtatatgat 692

Db    174 CCTCTCCACCTGCCAGGGGCTGCATGGGGTCTTCTCTGATGGGAACTTGACTTACATCGT 233

Qy    693 agagccactagagctggttcatgatgagaaaagcac------aggtcgaccacatataat 746

Db    234 GGAGCCCCAAGAGGTGGCTGGACCTTGGGGAGCCCCTCAGGGACCCCTTCCCCACCTCAT 293

Qy    747 ccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcactatggaaagagg 806

Db    294 TTACCGGACCCCTCTCCTCCCAGATCCCCTCGGATGCAGGGAACCAGGCTGCCTGTTTGC 353

Qy    807 tgaccagtggccctttctctctgaattacagtggttgaaaagaaggaagagagcagtgaa 866

Db    354 TGTGCCTGCCCAGTCGGCTCCTCCAAACCGGCCGAGGCTGAGAAGGAAAAGGCAGGTCCG 413

Qy    867 tccatcacgtggtatatttga---agaaatgaaatatttggaacttatgattgttaatga 923

Db    414 CCGGGGCCACCCTACAGTGCACAGTGAAACCAAGTATGTGGAGCTAATTGTGATCAACGA 473

Qy    924 tcacaaaacgtataagaagcatcgctcttctcatgcacataccaacaactttgcaaagtc 983

Db    474 CCACCAGCTGTTCGAGCAGATGCGACAGTCGGTGGTCCTCACCAGCAACTTTGCCAAGTC 533

Qy    984 cgtggtcaaccttgtggattctatttacaaggagcagctcaacaccagggttgtcctggt 1043

Db    534 CGTGGTGAACCTGGCCGATGTGATATACAAGGAGCAGCTCAACACTCGCATCGTCCTGGT 593

Qy   1044 ggctgtagagacctggactgagaaggatcagattgacatcaccaccaaccctgtgcagat 1103

Db    594 TGCCATGGAAACATGGGCAGATGGGGACAAGATCCAGGTGCAGGATGACCTCCTGGAGAC 653

Qy   1104 gctccatgagttctcaaaatac---cggcagcgcattaagcagcatgctgatgctgtgca 1160

Db    654 CCTGGCCCGGCTCATGGTCTACCGACGGGAGGGTCTGCCTGAGCCCAGTAATGCCACCCA 713

Qy   1161 cctcatctcgcgggtgacatttcactataagagaagcagtctgagttactttggaggtgt 1220

Db    714 CCTCTTCTCGGGCAGGACCTTCCAGAGCACGAGCAGCGGGGCAGCCTACGTGGGGGGCAT 773

Qy   1221 ctgttctcgcacaagaggagttggtgtgaatgagtatggtcttccaatggcagtggcaca 1280
```

```
Db    774 ATGCTCCCTGTCCCATGGCGGGGGTGTGAACGAGTACGGCAACATGGGGGCGATGGCCGT 833

Qy   1281 agtattatcgcagagcctggctcaaaaccttggaatccaatgggaaccttctagcagaaa 1340
Db    834 GACCCTTGCCCAGACGCTGGGACAGAACCTGGGCATGATGTGGAACAAACACCGGAGCTC 893

Qy   1341 gcca---aaatgtgactgcacagaatcctggggtggctgcatcatggaggaaacaggggt 1397
Db    894 GGCAGGGGACTGCAAGTGTCCAGACATCTGGCTGGGCTGCATCATGGAGGACACTGGGTT 953

Qy   1398 gtcccattctcgaaaattttcaaagtgcagcattttggagtatagagacttttttacagag 1457
Db    954 CTACCTGCCCCCGCAAGTTCTCTCGCTGCAGCATCGACGAGTACAACCAGTTTCTGCAGGA 1013

Qy   1458 aggaggtggagcctgccttttcaacaggccaacaaagctatttgagcccacggaatgtgg 1517
Db   1014 GGGTGGTGGCAGCTGCCTCTTCAACAAGCCCCTCAAGCTCCTGGACCCCCCAGAGTGCGG 1073

Qy   1518 aaatggatacgtggaagctggggaggagtgtgattgtggttttcatgtggaatgc----- 1572
     1074 GAACGGCTTCGTGGAGGCAGGGGAGGAGTGCGACTGCGGCTCGGTGCAGGAGTGCAGCCG 1133

     1573 ----tatggattatgctgtaagaaatgttccctctccaacggggctcactgcagcgacgg 1628
Db   1134 CGCAGGTGGCAACTGCTGCAAGAAATGCACCCTGACTCACGACGCCATGTGCAGCGACGG 1193

Qy   1629 gccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgccgggatgc 1688
Db   1194 GCTCTGCTGTCGCCGC------TGCAAGTACGAACCACGGGGTGTGTCCTGCCGAGAGGC 1247

Qy   1689 tgtgaacgagtgtgatattactgaatattgtactggagactctggtcagtgcccaccaaa 1748
Db   1248 CGTGAACGAGTGCGACATCGCGGAGACCTGCACCGGGGACTCTAGCCAGTGCCCGCCTAA 1307

Qy   1749 tcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctacaatggcga 1808
Db   1308 CCTGCACAAGCTGGACGGTTACTACTGTGACCATGAGCAGGGCCGCTGCTACGGAGGTCG 1367

Qy   1809 gtgcaagaccagagacaaccagtgtcagtacatctggggaacaaaggctgcagggtctga 1868
Db   1368 CTGCAAAACCCGGGACCGGCAGTGCCAGGTTCTTTGGGGCCATGCGGCTG------CTGA 1421

Qy   1869 caagttctgctatgaaaagctgaatacagaaggcactgagaagggaaactgcgggaagga 1928
Db   1422 TCGCTTCTGCTACGAGAAGCTGAATGTGGAGGGGACGGAGCGTGGGAGCTGTGGGCGCAA 1481

Qy   1929 tggagaccggtggattcagtgcagcaaaca----------tgatgtgttctgtggattct 1978
Db   1482 GGGATCCGGCTGGGTCCAGTGCAGTAAGCAGCCCCAACAGGGACGTGCTGTGTGGCTTCC 1541

     1979 tactctgtaccaatcttactcgagctccacgtattggtcaacttcagggtgagatcattc 2038
Db   1542 TCCTCTGTGTCAACATCTCTGGAGCTCCTCGGCTAGGGGACCTGGTGGGAGACATCAGTA 1601

Qy   2039 caacttccttctaccatcaaggccgggtgattgactgcagtggtgcccatgtagttttag 2098
Db   1602 GTGTCACCTTCTACCACCAGGGCAAGGAGCTGGACTGCAGGGGAGGCCACGTGCAGCTGG 1661

Qy   2099 atgatgatacggatgtgggctatgtagaagatggaacgccatgtggcccgtctatgatgt 2158
Db   1662 CGGACGGCTCTGACCTGAGCTATGTGGAGGATGGCACAGCCTGCGGGCCTAACATGTTGT 1721

Qy   2159 gtttagatcggaagtgcctacaaattcaagccctaaatatgagcagctgtccactcgatt 2218
Db   1722 GCCTGGACCATCGCTGCCTGCCAGCTTCTGCCTTCAACTTCAGCACCTGCCCCGGCAGTG 1781

Qy   2219 ccaagggtaaagtctgttcgggccatggggtgtgtagtaatgaagccacctgcatttgtg 2278
Db   1782 GGGAGCGCCGGATTTGCTCCCACCACGGGGTCTGCAGCAATGAAGGGAAGTGCATCTGTC 1841

Qy   2279 atttcacctgggcagggacagattgcagtatccgggatccagt---taggaaccttcacc 2335

Db   1842 AGCCAGACTGGACAGGCAAAGACTGCAGTATCCATAACCCCCTGCCCACGTCCCCACCCA 1901

Qy   2336 cccccaaggatgaaggacccaagggtcctagtgccaccaatctcataataggctccatcg 2395
Db   1902 CGGGGGAGACGGAGAGATATAAAGGTCCCAGCGGCACCAACATCATCATTGGCTCCATCG 1961

Qy   2396 ctggtgccatcctggtagcagctattgtccttgggggcacaggctggggatttaaaaatg 2455
Db   1962 CTGGGGCTGTCCTGGTTGCAGCCATCGTCCTGGGCGGCACGGGCTGGGGATTTAAAAACA 2021

Qy   2456 tcaagaagagaaggttcg 2473
Db   2022 TTCGCCGAGGAAGGTCCG 2039
```

```
RESULT 10
US-08-243-542-5
; Sequence 5, Application US/08243542
; Patent No. 5552526
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  :  EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/243,542
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
;      TELEFAX:  (616) 381-5465
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1464 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY:  human fetal brain cDNA library
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1464
```

US-08-243-542-5

Query Match 13.5%; Score 338.4; DB 1; Length 1464;
Best Local Similarity 56.2%; Pred. No. 3.7e-85;
Matches 793; Conservative 0; Mismatches 581; Indels 36; Gaps 7;

```
Qy    573 gggtggagagcactgttactaccatggaagcatcagaggcgtcaaacactccaaggtggc 632
Db     66 GGCTGGAGACCACTGCTACTACCAGGGGAAGCTCCGGGGGAACCCGCACTCCTTCGCCGC 125

Qy    633 tctgtcaacctgcaatggacttcatggcatgtttgaagatgataccttcgtgtatatgat 692
Db    126 CCTCTCCACCTGCCAGGGGCTGCATGGGGTCTTCTCTGATGGGAACTTGACTTACATCGT 185

Qy    693 agagccactagagctggttcatgatgagaaaagcac------aggtcgaccacatataat 746
Db    186 GGAGCCCCAAGAGGTGGCTGGACCTTGGGGAGCCCCTCAGGGACCCCTTCCCCACCTCAT 245

Qy    747 ccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcactatggaaagagg 806
Db    246 TTACCGGACCCCTCTCCTCCCAGATCCCCTCGGATGCAGGGAACCAGGCTGCCTGTTTGC 305

Qy    807 tgaccagtggccctttctctctgaattacagtggttgaaaagaaggaagagagcagtgaa 866
Db    306 TGTGCCTGCCCAGTCGGCTCCTCCAAACCGGCCGAGGCTGAGAAGGAAAAGGCAGGTCCG 365

Qy    867 tccatcacgtggtatatttga---agaaatgaaatatttggaacttatgattgttaatga 923
Db    366 CCGGGGCCACCCTACAGTGCACAGTGAAACCAAGTATGTGGAGCTAATTGTGATCAACGA 425

Qy    924 tcacaaaacgtataagaagcatcgctcttctcatgcacataccaacaactttgcaaagtc 983
Db    426 CCACCAGCTGTTCGAGCAGATGCGACAGTCGGTGGTCCTCACCAGCAACTTTGCCAAGTC 485

Qy    984 cgtggtcaaccttgtggattctatttacaaggagcagctcaacaccagggttgtcctggt 1043
Db    486 CGTGGTGAACCTGGCCGATGTGATATACAAGGAGCAGCTCAACACTCGCATCGTCCTGGT 545

Qy   1044 ggctgtagagacctggactgagaaggatcagattgacatcaccaccaaccctgtgcagat 1103
Db    546 TGCCATGGAAACATGGGCAGATGGGGACAAGATCCAGGTGCAGGATGACCTCCTGGAGAC 605

Qy   1104 gctccatgagttctcaaaatac---cggcagcgcattaagcagcatgctgatgctgtgca 1160
Db    606 CCTGGCCCGGCTCATGGTCTACCGACGGGAGGGTCTGCCTGAGCCCAGTAATGCCACCCA 665

Qy   1161 cctcatctcgcgggtgacatttcactataagagaagcagtctgagttacttttggaggtgt 1220
Db    666 CCTCTTCTCGGGCAGGACCTTCCAGAGCACGAGCAGCGGGGCAGCCTACGTGGGGGGCAT 725

Qy   1221 ctgttctcgcacaagaggagttggtgtgaatgagtatggtcttccaatggcagtggcaca 1280
Db    726 ATGCTCCCTGTCCCATGGCGGGGGTGTGAACGAGTACGGCAACATGGGGGCGATGGCCGT 785

Qy   1281 agtattatcgcagagcctggctcaaaaccttggaatccaatgggaacctttctagcagaaa 1340
Db    786 GACCCTTGCCCAGACGCTGGGACAGAACCTGGGCATGATGTGGAACAAACACCGGAGCTC 845

Qy   1341 gcca---aaatgtgactgcacagaatcctggggtggctgcatcatgggaggaaacaggggt 1397
Db    846 GGCAGGGGACTGCAAGTGTCCAGACATCTGGCTGGGCTGCATCATGGAGGACACTGGGTT 905

Qy   1398 gtcccattctcgaaaattttcaaagtgcagcattttggagtatagagacttttttacagag 1457
Db    906 CTACCTGCCCCGCAAGTTCTCTCGCTGCAGCATCGACGAGTACAACCAGTTTCTGCAGGA 965

Qy   1458 aggaggtggagcctgccttttcaacaggccaacaaagctatttgagccacggaatgtgg 1517
Db    966 GGGTGGTGGCAGCTGCCTCTTCAACAAGCCCCTCAAGCTCCTGGACCCCCCAGAGTGCGG 1025

Qy   1518 aaatggatacgtggaagctggggaggagtgtgattgtggttttcatggaatgc----- 1572
Db   1026 GAACGGCTTCGTGGAGGCAGGGGAGGAGTGCGACTGCGGCTCGGTGCAGGAGTGCAGCCG 1085

Qy   1573 ----tatggattatgctgtaagaaaatgttccctctccaacggggctcactgcagcgacgg 1628
Db   1086 CGCAGGTGGCAACTGCTGCAAGAAATGCACCCTGACTCACGACGCCATGTGCAGCGACGG 1145

Qy   1629 gccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgccgggatgc 1688
Db   1146 GCTCTGCTG------TCGCCGCTGCAAGTACGAACCACGGGGTGTGTCCTGCCGAGAGGC 1199

Qy   1689 tgtgaacgagtgtgatattactgaatattgtactggagactctggtcagtgcccaccaaa 1748
Db   1200 CGTGAACGAGTGCGACATCGCGGAGACCTGCACCGGGGACTCTAGCCAGTGCCCGCCTAA 1259

Qy   1749 tcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctacaatggcga 1808
Db   1260 CCTGCACAAGCTGGACGGTTACTACTGTGACCATGAGCAGGGCCGCTGCTACGGAGGTCG 1319

Qy   1809 gtgcaagaccagagacaaccagtgtcagtacatctggggaacaaaggctgcagggtctga 1868
Db   1320 CTGCAAAACCCGGGACCGGCAGTGCCAGGTTCTTTGGGGCCATGCGGCTG------CTGA 1373

Qy   1869 caagttctgctatgaaaagctgaatacagaaggcactgagaagggaaactgcgggaagga 1928
Db   1374 TCGCTTCTGCTACGAGAAGCTGAATGTGGAGGGGACGGAGCGTGGGAGCTGTGGGCGCAA 1433

Qy   1929 tggagaccggtggattcagtgcagcaaaca 1958
Db   1434 GGGATCCGGCTGGGTCCAGTGCAGTAAGCA 1463
```

RESULT 11
US-08-477-407-5
; Sequence 5, Application US/08477407
; Patent No. 5631351
; GENERAL INFORMATION:
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: : EMI, MITSURU
; TITLE OF INVENTION: MDC PROTEINS AND DNAs
; TITLE OF INVENTION: ENCODING THE SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLYNN, THIEL, BOUTELL & TANIS P.C.
; STREET: 2026 Rambling Road
; CITY: Kalamazoo
; STATE: Michigan
; COUNTRY: USA
; ZIP: 49008-1699
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage
; COMPUTER: IBM PC/XT/AT Compatible
; OPERATING SYSTEM: MS-DOS 5.0
; SOFTWARE: WordPerfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,407
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/243,542
; FILING DATE: 13-MAY-1994
; APPLICATION NUMBER: JP 5-136602
; FILING DATE: 14 MAY 1993
; APPLICATION NUMBER: JP 5-257455
; FILING DATE: 22 SEPTEMBER 1993
; APPLICATION NUMBER: JP 6-49904
; FILING DATE: 23 FEBRUARY 1994
; APPLICATION NUMBER: JP 6-73328
; FILING DATE: 12 APRIL 1994
; APPLICATION NUMBER: JP 6-84470
; FILING DATE: 22 APRIL 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Terryence F. Chapman
; REGISTRATION NUMBER: 32 549

```
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
;      TELEFAX:  (616) 381-5465
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1464 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY:  human fetal brain cDNA library
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1464
US-08-477-407-5

[redacted]ery Match            13.5%;  Score 338.4;  DB 1;  Length 1464;
[redacted]st Local Similarity   56.2%;  Pred. No. 3.7e-85;
[redacted]tches  793;  Conservative    0;  Mismatches  581;  Indels   36;  Gaps    7;

Qy    573 gggtggagagcactgttactaccatggaagcatcagaggcgtcaaagactccaaggtggc 632
          || |||||| ||||| |||||||| || |   || | ||   |    |||||   |  ||
Db     66 GGCTGGAGACCACTGCTACTACCAGGGGAAGCTCCGGGGGAACCCGCACTCCTTCGCCGC 125

Qy    633 tctgtcaacctgcaatggacttcatggcatgtttgaagatgataccttcgtgtatatgat 692
          || || |||||| | || || |||||  | ||    ||||  | |||    || ||  |
Db    126 CCTCTCCACCTGCCAGGGGCTGCATGGGGTCTTCTCTGATGGGAACTTGACTTACATCGT 185

Qy    693 agagccactagagctggttcatgatgagaaaagcac------aggtcgaccacatataat 746
          ||||| | |||| ||| |     |  |   | | |       |   |  || ||  | ||
Db    186 GGAGCCCCAAGAGGTGGCTGGACCTTGGGGAGCCCCTCAGGGACCCCTTCCCCACCTCAT 245

Qy    747 ccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcactatggaaagagg 806
            |    |||         |   || |      |   | | | |              |
Db    246 TTACCGGACCCCTCTCCTCCCAGATCCCCTCGGATGCAGGGAACCAGGCTGCCTGTTTGC 305

Qy    807 tgaccagtggccctttctctctgaattacagtggttgaaaagaaggaagagagcagtgaa 866
          ||  |     |  |       ||  |   | | |  |   |||||||| ||     ||
Db    306 TGTGCCTGCCCAGTCGGCTCCTCCAAACCGGCCGAGGCTGAGAAGGAAAAGGCAGGTCCG 365

Qy    867 tccatcacgtggtatatttga---agaaatgaaatatttggaacttatgattgttaatga 923
          |     |    || | | |    ||||  || ||| |||| || ||  |   | || ||
Db    366 CCGGGGCCACCCTACAGTGCACAGTGAAACCAAGTATGTGGAGCTAATTGTGATCAACGA 425

Qy    924 tcacaaaacgtataagaagcatcgctcttctcatgcacataccaacaactttgcaaagtc 983
          ||| |   ||   || ||   ||    ||    |  |  |||| |||||||| |||||
Db    426 CCACCAGCTGTTCGAGCAGATGCGACAGTCGGTGGTCCTCACCAGCAACTTTGCCAAGTC 485

Qy    984 cgtggtcaaccttgtggattctatttacaaggagcagctcaacaccagggttgtcctggt 1043
          |||||| ||||| |  |||   || ||||||||||||||||||  |  | ||||||||
Db    486 CGTGGTGAACCTGGCCGATGTGATATACAAGGAGCAGCTCAACACTCGCATCGTCCTGGT 545

Qy   1044 ggctgtagagacctggactgagaaggatcagattgacatcaccaccaaccctgtgcagat 1103
          ||  | || || ||| | ||   |||  |||| |  |        |||  || |||
Db    546 TGCCATGGAAACATGGGCAGATGGGGACAAGATCCAGGTGCAGGATGACCTCCTGGAGAC 605

Qy   1104 gctccatgagttctcaaaatac---cggcagcgcattaagcagcatgctgatgctgtgca 1160
          ||      | ||      |||   ||| || |  |     |||   . | ||||    ||
Db    606 CCTGGCCCGGCTCATGGTCTACCGACGGGAGGGTCTGCCTGAGCCCAGTAATGCCACCCA 665

Qy   1161 cctcatctcgcgggtgacatttcactataagagaagcagtctgagttactttggaggtgt 1220
          |||| ||||| |   ||| || ||    | ||| ||| |      ||| | || ||  |
Db    666 CCTCTTCTCGGGCAGGACCTTCCAGAGCACGAGCAGCGGGGCAGCCTACGTGGGGGGCAT 725

Qy   1221 ctgttctcgcacaagaggagttggtgtgaatgagtatggtcttccaatggcagtggcaca 1280
          || || |   |    || |  |||||||| ||||| ||          |||  ||||
Db    726 ATGCTCCCTGTCCCATGGCGGGGGTGTGAACGAGTACGGCAACATGGGGGCGATGGCCGT 785

Qy   1281 agtattatcgcagagcctggctcaaaaccttggaatccaatgggaaccttctagcagaaa 1340
               |  | ||||  ||||  || ||||| || ||    ||| |        | ||
Db    786 GACCCTTGCCCAGACGCTGGGACAGAACCTGGGCATGATGTGGAACAAACACCGGAGCTC 845

Qy   1341 gcca---aaatgtgactgcacagaatcctggggtggctgcatcatggaggaaaacaggggt 1397
          | ||    | ||  | ||  ||||   ||||   ||||||||||||||||| || ||| |
Db    846 GGCAGGGGACTGCAAGTGTCCAGACATCTGGCTGGGCTGCATCATGGAGGACACTGGGTT 905

Qy   1398 gtcccattctcgaaaattttcaaagtgcagcattttggagtatagagacttttttacagag 1457
          | ||    | || || || ||    ||||||||    |||||  |   | ||| | |||
Db    906 CTACCTGCCCCGCAAGTTCTCTCGCTGCAGCATCGACGAGTACAACCAGTTTCTGCAGGA 965

Qy   1458 aggaggtggagcctgccttttcaacaggccaacaaagctatttgagcccacggaatgtgg 1517
          || |||||   |||||| |||||||| |||    |||||  | || ||| | || || ||
Db    966 GGGTGGTGGCAGCTGCCTCTTCAACAAGCCCCTCAAGCTCCTGGACCCCCCAGAGTGCGG 1025

Qy   1518 aaatggatacgtggaagctggggaggagtgtgattgtggttttcatgtggaatgc----- 1572
          || || | |||||| || |||||||||| || || || |       ||| |||
Db    1026 GAACGGCTTCGTGGAGGCAGGGGAGGAGTGCGACTGCGGCTCGGTGCAGGAGTGCAGCCG 1085

Qy   1573 ----tatggattatgctgtaagaaatgttccctctccaacggggctcactgcagcgacgg 1628
                 |||    ||||| ||||||||  ||||  |  |||  ||    ||||||||||
Db    1086 CGCAGGTGGCAACTGCTGCAAGAAATGCACCCTGACTCACGACGCCATGTGCAGCGACGG 1145

Qy   1629 gccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgccgggatgc 1688
          || ||||||      |  |   ||     |   | ||||| ||       ||||| || ||
Db    1146 GCTCTGCTG------TCGCCGCTGCAAGTACGAACCACGGGGTGTGTCCTGCCGAGAGGC 1199

Qy   1689 tgtgaacgagtgtgatattactgaatattgtactggagactctggtcagtgcccaccaaa 1748
          |||||||||||| || ||  | ||    || || || |||||| | |||||||| || ||
Db    1200 CGTGAACGAGTGCGACATCGCGGAGACCTGCACCGGGGACTCTAGCCAGTGCCCGCCTAA 1259

Qy   1749 tcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctacaatggcga 1808
          || || ||||  |||||  ||    ||  | ||  | ||||||||||||||   ||
Db    1260 CCTGCACAAGCTGGACGGTTACTACTGTGACCATGAGCAGGGCCGCTGCTACGGAGGTCG 1319

Qy   1809 gtgcaagaccagagacaaccagtgtcagtacatctggggaacaaaggctgcagggtctga 1868
          ||||| ||| | |||   |||||  |||    | |||||      |||||      ||||
Db    1320 CTGCAAAACCCGGGACCGGCAGTGCCAGGTTCTTTGGGGCCATGCGGCTG------CTGA 1373

Qy   1869 caagttctgctatgaaaagctgaatacagaaggcactgagaagggaaactgcgggaagga 1928
          ||||||||  || |||||||||   || || || |||   || | ||| |||    |
Db    1374 TCGCTTCTGCTACGAGAAGCTGAATGTGGAGGGGACGGAGCGTGGGAGCTGTGGGCGCAA 1433

Qy   1929 tggagaccggtggattcagtgcagcaaaca 1958
          |||  | | ||| | |||||||| || ||
Db    1434 GGGATCCGGCTGGGTCCAGTGCAGTAAGCA 1463


RESULT  12
US-08-484-355-5
; Sequence 5, Application US/08484355
; Patent No. 5705341
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  :  EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,355
```

```
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/243,542
;      FILING DATE:  13-MAY-1994
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
;      TELEFAX:  (616) 381-5465
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1464 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY:  human fetal brain cDNA library
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1464
US-08-484-355-5


  Query Match             13.5%;  Score 338.4;  DB 1;  Length 1464;
  Best Local Similarity   56.2%;  Pred. No. 3.7e-85;
  Matches  793;  Conservative    0;  Mismatches  581;  Indels   36;  Gaps    7;

Qy    573 gggtggagagcactgttactaccatggaagcatcagaggcgtcaaagactccaaggtggc 632
          || |||||| ||||| |||||||| || |   || | ||   |    |||||   |  ||
Db     66 GGCTGGAGACCACTGCTACTACCAGGGGAAGCTCCGGGGGAACCCGCACTCCTTCGCCGC 125

Qy    633 tctgtcaacctgcaatggacttcatggcatgtttgaagatgataccttcgtgtatatgat 692
          || || |||||| | || || |||||  | ||    ||||  | |||  || ||  |
Db    126 CCTCTCCACCTGCCAGGGGCTGCATGGGGTCTTCTCTGATGGGAACTTGACTTACATCGT 185

Qy    693 agagccactagagctggttcatgatgagaaaagcac------aggtcgaccacatataat 746
          |||||| | |||| ||| |      |  |  | | |       |   |  || ||  | ||
Db    186 GGAGCCCCAAGAGGTGGCTGGACCTTGGGGAGCCCCTCAGGGACCCCTTCCCCACCTCAT 245

Qy    747 ccagaaaaccttggcaggacagtattctaagcaaatgaagaatctcactatggaaagagg 806
            |   |||          |   || |      |   | | | |              |
Db    246 TTACCGGACCCCTCTCCTCCCAGATCCCCTCGGATGCAGGGAACCAGGCTGCCTGTTTGC 305

Qy    807 tgaccagtggccctttctctctgaattacagtggttgaaaagaaggaagagagcagtgaa 866
          ||  |   |  |      ||  |   | |  |  |  |||||||| ||    ||
Db    306 TGTGCCTGCCCAGTCGGCTCCTCCAAACCGGCCGAGGCTGAGAAGGAAAAGGCAGGTCCG 365

Qy    867 tccatcacgtggtatatttga---agaaatgaaatatttggaacttatgattgttaatga 923
          |     |    || | |  |    ||||  || ||| |||| || ||   |  | || ||
Db    366 CCGGGGCCACCCTACAGTGCACAGTGAAACCAAGTATGTGGAGCTAATTGTGATCAACGA 425

Qy    924 tcacaaaacgtataagaagcatcgctcttctcatgcacataccaacaacttttgcaaagtc 983
          ||| |   ||  || ||   ||    ||     |  |  |||| |||||||| |||||
Db    426 CCACCAGCTGTTCGAGCAGATGCGACAGTCGGTGGTCCTCACCAGCAACTTTGCCAAGTC 485

Qy    984 cgtggtcaaccttgtggattctatttacaaggagcagctcaacaccagggttgtcctggt 1043
          |||||| ||||| |  |||   || ||||||||||||||||||||  |    ||||||||

Db    486 CGTGGTGAACCTGGCCGATGTGATATACAAGGAGCAGCTCAACACTCGCATCGTCCTGGT 545

Qy   1044 ggctgtagagacctggactgagaaggatcagattgacatcaccaccaaccctgtgcagat 1103
          ||  | || || ||| | ||   |||  ||||  |  |         |||  || |||
Db    546 TGCCATGGAAACATGGGCAGATGGGGACAAGATCCAGGTGCAGGATGACCTCCTGGAGAC 605

Qy   1104 gctccatgagttctcaaaatac---cggcagcgcattaagcagcatgctgatgctgtgca 1160
          ||       | ||       |||   ||| || |  |      |||    | ||||    ||
Db    606 CCTGGCCCGGCTCATGGTCTACCGACGGGAGGGTCTGCCTGAGCCCAGTAATGCCACCCA 665

Qy   1161 cctcatctcgcgggtgacatttcactataagagaagcagtctgagttactttggaggtgt 1220
          |||| ||||| |   ||| || ||    | ||| ||| |        ||| | || ||  |
Db    666 CCTCTTCTCGGGCAGGACCTTCCAGAGCACGAGCAGCGGGGCAGCCTACGTGGGGGGCAT 725

Qy   1221 ctgttctcgcacaagaggagttggtgtgaatgagtatggtcttccaatggcagtggcaca 1280
          || || |   |    || |  |||||||| ||||| ||        . |||  ||||
Db    726 ATGCTCCCTGTCCCATGGCGGGGGTGTGAACGAGTACGGCAACATGGGGGCGATGGCCGT 785

Qy   1281 agtattatcgcagagcctggctcaaaaccttggaatccaatgggaaccttctagcagaaa 1340
               |  | ||||  ||||  || ||||| || ||    |||  |        | ||
Db    786 GACCCTTGCCCAGACGCTGGGACAGAACCTGGGCATGATGTGGAACAAACACCGGAGCTC 845

Qy   1341 gcca---aaatgtgactgcacagaatcctggggtggctgcatcatggaggaaacaggggt 1397
          | ||    | ||  | ||  ||||   ||||   |||||||||||||||||| || ||| |
Db    846 GGCAGGGGACTGCAAGTGTCCAGACATCTGGCTGGGCTGCATCATGGAGGACACTGGGTT 905

Qy   1398 gtcccattctcgaaaattttcaaagtgcagcattttggagtatagagacttttacagag 1457
          | ||   | || || || ||    |||||||   |||||  |  | ||| | |||
Db    906 CTACCTGCCCCGCAAGTTCTCTCGCTGCAGCATCGACGAGTACAACCAGTTTCTGCAGGA 965

Qy   1458 aggaggtggagcctgccttttcaacaggccaacaaagctatttgagcccacggaatgtgg 1517
          || |||||   |||||| ||||||| |||    |||||  | || ||| | || || ||
Db    966 GGGTGGTGGCAGCTGCCTCTTCAACAAGCCCCTCAAGCTCCTGGACCCCCCAGAGTGCGG 1025

Qy   1518 aaatggatacgtggaagctggggaggagtgtgattgtggttttcatgtggaatgc----- 1572
          || || | |||||| || |||||||||||| || || || |       ||| |||
Db   1026 GAACGGCTTCGTGGAGGCAGGGGAGGAGTGCGACTGCGGCTCGGTGCAGGAGTGCAGCCG 1085

Qy   1573 ----tatggattatgctgtaagaaatgttccctctccaacggggctcactgcagcgacgg 1628
                 |||    ||||| ||||||||  ||||  |  |||  ||    ||||||||||
Db   1086 CGCAGGTGGCAACTGCTGCAAGAAATGCACCCTGACTCACGACGCCATGTGCAGCGACGG 1145

Qy   1629 gccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgccgggatgc 1688
          || ||||||      |  |   ||     |  | ||||| ||       ||||| || ||
Db   1146 GCTCTGCTG------TCGCCGCTGCAAGTACGAACCACGGGGTGTGTCCTGCCGAGAGGC 1199

Qy   1689 tgtgaacgagtgtgatattactgaatattgtactggagactctggtcagtgcccaccaaa 1748
          |||||||||| || ||  | ||    || || || |||||| | |||||||| || ||
Db   1200 CGTGAACGAGTGCGACATCGCGGAGACCTGCACCGGGGACTCTAGCCAGTGCCCGCCTAA 1259

Qy   1749 tcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctacaatggcga 1808
          ||  || ||||  |||||  ||    ||  | ||  | ||||||||||||||   ||
Db   1260 CCTGCACAAGCTGGACGGTTACTACTGTGACCATGAGCAGGGCCGCTGCTACGGAGGTCG 1319

Qy   1809 gtgcaagaccagagacaaccagtgtcagtacatctggggaacaaaggctgcagggtctga 1868
          |||||  ||| | |||   |||||  |||    | |||||      |||||      ||||
Db   1320 CTGCAAAACCCGGGACCGGCAGTGCCAGGTTCTTTGGGGCCATGCGGCTG------CTGA 1373

Qy   1869 caagttctgctatgaaaagctgaatacagaaggcactgagaagggaaactgcgggaagga 1928
             ||||||||  || |||||||||   || || || |||   || | ||| |||    |
Db   1374 TCGCTTCTGCTACGAGAAGCTGAATGTGGAGGGGACGGAGCGTGGGAGCTGTGGGCGCAA 1433

Qy   1929 tggagaccggtggattcagtgcagcaaaca 1958
          |||  | | ||| | |||||||| || ||
Db   1434 GGGATCCGGCTGGGTCCAGTGCAGTAAGCA 1463


RESULT  13
US-08-836-443-1
; Sequence 1, Application US/08836443
; Patent No. 5883241
;  GENERAL INFORMATION:
;    APPLICANT:  DOCHERTY, Andrew, J.P
```

```
;    APPLICANT:  SLOCOMBE, Patrick, M.
;    TITLE OF INVENTION:  DNA SEQUENCES CODING FOR A
;    TITLE OF INVENTION:  HUMAN METALLOPROTEINASE AND VARIANTS THEREOF
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,443
;      FILING DATE:  01-MAY-1997
;      CLASSIFICATION:  435
     PRIOR APPLICATION DATA:
       APPLICATION NUMBER:  9709420
       FILING DATE:  05-SEP-1997
       APPLICATION NUMBER:  9612145.4
;      FILING DATE:  11-JUN-1996
;      APPLICATION NUMBER:  9526230.9
;      FILING DATE:  21-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Resnick, David S
;      REGISTRATION NUMBER:  34,235
;      REFERENCE/DOCKET NUMBER:  47424
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2648 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-836-443-1


  Query Match             4.3%;  Score 107.6;  DB 2;  Length 2648;
  Best Local Similarity   45.2%;  Pred. No. 2.7e-20;
  Matches  650;  Conservative    0;  Mismatches  764;  Indels   24;  Gaps    6;

Qy     892 atgaaatatttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctct 951
           ||||| ||| |||| |||    | ||   |||| |   |  || | ||||| ||||
Db     102 ATGAAGTATGTGGAGCTTTACCTCGTGGCTGATTATTTAGAGTTTCAGAAGAATCGACGA 161

       952 tctcatgcacataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttac 1011
              || |     |||||  ||      |      | | ||||  ||| |||    |||||
       162 GACCAGGACGCCACCAAACACAAGCTCATAGAGATCGCCAACTATGTTGATAAGTTTTAC 221

Qy    1012 aaggagcagctcaacaccagggttgtcctgg---tggctgtagagacctggactgagaag 1068
                 |  || |   | | | || |||   |||  ||   ||||       | | |
Db     222 CGATCCTTGAACATCCGGATTGCTCTCGTGGGCTTGGAAGTGTGGACCCACGGGAACATG 281

Qy    1069 gatcagattgacatcaccaccaaccctgtgcagatgctccatgagttctcaaaataccgg 1128
             | |  ||      |   | |  ||   |    | | |       | |    |  |
Db     282 TGTGAAGTTTCAGAGAATCCATATTCTACCCTCTGGTCCTTTCTCAGTTGGAGGCGCAAG 341

Qy    1129 cagcgcattaagcagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactat 1188
           | ||      || || |   |||    |  ||  | ||| || |  || | || |||
Db     342 CTGCTTGCCCAGAAGTACCATGACAACGCCCAATTAATCACGGGCATGTCCTTCCACGGC 401

Qy    1189 aagagaagcagtctgagttactttggaggtgtctgttct------cgcacaagaggagtt 1242
           |  |  | | |||     | |    |   | || |||       |   |  |||||||
Db     402 ACCACCATCGGCCTGGCCCCCCTCATGGCCATGTGCTCTGTGTACCAGTCTGGAGGAGTC 461

Qy    1243 ggtgtgaatgagtatggtcttccaatggcagtggcacaagtattatcgcagagcctggct 1302
              || |  | |  |    | | || |  |||||        |  | ||     |||
Db     462 AACATGGACCACTCCGAGAATGCCATTGGCGTGGCTGCCACCATGGCCCACGAGATGGGC 521

Qy    1303 caaaaccttggaatccaatgggaacctttctagcagaaagccaaaatgtgactgcacagaa 1362
           || ||| |||| ||       ||  || |      | |       |  | | | | |
Db     522 CACAACTTTGGCATGACCCATGATTCTGC------AGATTGCTGCTCGGCCAGTGCGGCT 575

Qy    1363 tcctggggtggctgcatcatggaggaaacaggggtgtcccattctcgaaaattttcaaag 1422
              || ||  |||   |  |  || |||    ||  | |    |   ||
Db     576 GATGGTGGGTGCATCATGGCAGCTGCCACTGGGCACCCCTTTCCCAAAGTGTTCAATGGA 635

Qy    1423 tgcagcattttggagtatagagacttttacagagaggaggtggagcctgccttttcaac 1482
           ||||  ||    ||||         | | | |||  ||| ||||||   || || | ||||
Db     636 TGCAACAGGAGGGAGCTGGACAGGTATCTGCAGTCAGGTGGTGGAATGTGTCTCTCCAAC 695

Qy    1483 aggcca------acaaagctatttgagcccacggaatgtggaaatggatacgtggaagct 1536
           | ||||      |  | | | | ||    |  |   ||||| || || ||  |||||| |
Db     696 ATGCCAGACACCAGGATGTTGTATGGAGGCCGGAGGTGTGGGAACGGGTATCTGGAAGAT 755

Qy    1537 ggggaggagtgtgattgtggttttcatgtggaatgctatggattatgctgtaagaaatgt 1596
           ||||| |||||||| |||||     | | ||||||  |       |||||  ||    | |
Db     756 GGGGAAGAGTGTGACTGTGGAGAAGAAGAGGAATGTAACAACCCCTGCTGCAATGCCTCT 815

Qy    1597 tccctctccaacggggctcactgcagcgacgggccctgctgtaacaatacctcatgtctt 1656
                ||    || |           |  | | |||    |      |   |||
Db     816 AATTGTACCCTGAGGCCGGGGGCGGAGTGTGCTCACGGCTCCTGCTGCCACCAGTGTAAG 875

Qy    1657 tttcagccacgagggtatgaatgccgggatgctgtgaacgagtgtgatattactgaatat 1716
             |  | | |  |||       ||||| ||     |  |  |||||||  |  | || |
Db     876 CTGTTGGCTCCTGGGACCCTGTGCCGCGAGCAGGCCAGGCAGTGTGACCTCCCGGAGTTC 935

Qy    1717 tgtactggagactctggtcagtgcccaccaaatcttcataagcaagacggatatgcatgc 1776
           ||||| ||  | |||   || |||||  | ||  |  |  ||   || ||     | ||
Db     936 TGTACGGGCAAGTCTCCCCACTGCCCTACCAACTTCTACCAGATGGATGGTACCCCCTGT 995

Qy    1777 aatcaaaatcagggccgctgctacaatggcgagtgcaagaccagagacaaccagtgtcag 1836
             |       |||| |  |||||||||| |||  ||||   |||    |  | ||||| |||
Db     996 GAGGGCGGCCAGGCCTACTGCTACAACGGCATGTGCCTCACCTACCAGGAGCAGTGCCAG 1055

Qy    1837 tacatctggggaacaaaggctgcagggtctgacaagttctgctatgaaaagctgaataca 1896
             |  | |||||| |    ||   |    |    | ||||||  || ||| |||||
Db     1056 CAGCTGTGGGGACCCGGAGCCCGACCTGCCCCTGACCTCTGCTTCGAGAAGGTGAATGTG 1115

Qy    1897 gaaggcactgagaagggaaactgcgggaaggatggagaccggtggattcagtgcagcaaa 1956
           | |||          |||||||| || |||||    ||  |||| |  |||       ||
Db     1116 GCAGGAGACACCTTTGGAAACTGTGGAAAGGA-CATGAATGGTGAACACAGGAAGTGCAA 1174

Qy    1957 catgatgtgttctgtggattcttactctgtaccaatcttactcgagctccacgtattggt 2016
           |||||    | |   | |  |                  |||  ||    |||
Db     1175 CATGAGAGATGCGAAGTGTGGGAAGATCCAGTGTCAGAGCTCTGAGGCCCGGCCCCTGGA 1234

Qy    2017 caacttcagggtgagatcattccaacttccttctaccatcaaggccgggtgattgactgc 2076
             |  | ||||    ||||   || |  ||  |    | ||  ||  ||| |  |||
Db     1235 GTCCAACGCGGTGC--CCATTGACACCACTATCATCATGAATGGGAGGCAGATCCAGTGC 1292

Qy    2077 agtggtgcccatgtagttttagatgatgatacggatgtgggctatgtagaagatggaacg 2136
             | ||  |||| ||      || |  |||   ||| |  | |    | ||    ||   |
Db     1293 CGGGGCACCCACGTCTACCGAGGTCCTGAGGAGGAGGGTGACATGCTGGACCCAGGGCTG 1352

Qy    2137 ccatgtggcccgtctatgatgtgtttagatcggaagtgcctacaaattcaagccctaaat 2196
                     | | |    | |   | |  |  ||| |  |    ||   |   ||
Db     1353 GTGATGACTGGAACCAAGTGTGGCTACAACCATATTTGCTTTGAGGGGCAGTGCAGGAAC 1412

Qy    2197 atgagcagctgtccactcgattccaagggtaaagtctgttcgggccatggggtgtgtagt 2256
             |   || | |  ||    || ||     ||  ||||||||||| ||||
Db     1413 ACCTCCTTCTTTGAAACTGAAGGCTGTGGGAAGAAGTGCAATGGCCATGGGGTCTGTAAC 1472

Qy    2257 aatgaagccacctgcatttgtgatttcacctgggcagggacagattgcagtatccggg 2314
           ||  |    | ||||   ||         ||||||  | |    ||||  |  | ||
Db     1473 AACAACCAGAACTGCCACTGCCTGCCGGGCTGGGCCCCGCCCTTCTGCAACACACCGG 1530

RESULT  14
```

```
US-08-836-443-2
; Sequence 2, Application US/08836443
; Patent No. 5883241
;  GENERAL INFORMATION:
;    APPLICANT:  DOCHERTY, Andrew, J.P
;    APPLICANT:  SLOCOMBE, Patrick, M.
;    TITLE OF INVENTION:  DNA SEQUENCES CODING FOR A
;    TITLE OF INVENTION:  HUMAN METALLOPROTEINASE AND VARIANTS THEREOF.
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
;      STREET:  130 Water Street
;      CITY:  Boston
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/836,443
;      FILING DATE:  01-MAY-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  9709420
;      FILING DATE:  05-SEP-1997
;      APPLICATION NUMBER:  9612145.4
;      FILING DATE:  11-JUN-1996
;      APPLICATION NUMBER:  9526230.9
;      FILING DATE:  21-DEC-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Resnick, David S
;      REGISTRATION NUMBER:  34,235
;      REFERENCE/DOCKET NUMBER:  47424
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-523-3400
;      TELEFAX:  617-523-6440
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2251 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    08-836-443-2
```

Query Match 4.0%; Score 99.6; DB 2; Length 2251;
Best Local Similarity 46.5%; Pred. No. 4.4e-18;
Matches 478; Conservative 0; Mismatches 529; Indels 21; Gaps 4;

```
Qy    892 atgaaatatttggaacttatgattgttaatgatcacaaaacgtataagaagcatcgctct 951
Db    100 ATGAAGTATGTGGAGCTTTACCTCGTGGCTGATTATTTAGAGTTTCAGAAGAATCGACGA 159

Qy    952 tctcatgcacataccaacaactttgcaaagtccgtggtcaaccttgtggattctatttac 1011
Db    160 GACCAGGACGCCACCAAACACAAGCTCATAGAGATCGCCAACTATGTTGATAAGTTTTAC 219

Qy   1012 aaggagcagctcaacaccagggttgtcctgg---tggctgtagagacctggactgagaag 1068
Db    220 CGATCCTTGAACATCCGGATTGCTCTCGTGGGCTTGGAAGTGTGGACCCACGGGAACATG 279

Qy   1069 gatcagattgacatcaccaccaaccctgtgcagatgctccatgagttctcaaaataccgg 1128
Db    280 TGTGAAGTTTCAGAGAATCCATATTCTACCCTCTGGTCCTTTCTCAGTTGGAGGCGCAAG 339

Qy   1129 cagcgcattaagcagcatgctgatgctgtgcacctcatctcgcgggtgacatttcactat 1188
Db    340 CTGCTTGCCCAGAAGTACCATGACAACGCCCAATTAATCACGGGCATGTCCTTCCACGGC 399

Qy   1189 aagagaagcagtctgagttactttggaggtgtctgttct------cgcacaagaggagtt 1242
Db    400 ACCACCATCGGCCTGGCCCCCCTCATGGCCATGTGCTCTGTGTACCAGTCTGGAGGAGTC 459

Qy   1243 ggtgtgaatgagtatggtcttccaatggcagtggcacaagtattatcgcagagcctggct 1302
Db    460 AACATGGACCACTCCGAGAATGCCATTGGCGTGGCTGCCACCATGGCCCACGAGATGGGC 519

Qy   1303 caaaaccttggaatccaatgggaaccttctagcagaaagccaaaatgtgactgcacagaa 1362
Db    520 CACAACTTTGGCATGACCCATGATTCTGC------AGATTGCTGCTCGGCCAGTGCGGCT 573

Qy   1363 tcctggggtggctgcatcatggaggaaacaggggtgtcccattctcgaaaattttcaaag 1422
Db    574 GATGGTGGGTGCATCATGGCAGCTGCCACTGGGCACCCCTTTCCCAAAGTGTTCAATGGA 633

Qy   1423 tgcagcattttggagtatagagacttttacagagaggaggtggagcctgccttttcaac 1482
Db    634 TGCAACAGGAGGGAGCTGGACAGGTATCTGCAGTCAGGTGGTGGAATGTGTCTCTCCAAC 693

Qy   1483 aggcca------acaaagctatttgagcccacggaatgtggaaatggatacgtggaagct 1536
Db    694 ATGCCAGACACCAGGATGTTGTATGGAGGCCGGAGGTGTGGGAACGGGTATCTGGAAGAT 753

Qy   1537 ggggaggagtgtgattgtggttttcatgtggaatgctatggattatgctgtaagaaatgt 1596
Db    754 GGGGAAGAGTGTGACTGTGGAGAAGAAGAGGAATGTAACAACCCCTGCTGCAATGCCTCT 813

Qy   1597 tccctctccaacggggctcactgcagcgacgggccctgctgtaacaatacctcatgtctt 1656
Db    814 AATTGTACCCTGAGGCCGGGGGCGGAGTGTGCTCACGGCTCCTGCTGCCACCAGTGTAAG 873

Qy   1657 tttcagccacgagggtatgaatgccgggatgctgtgaacgagtgtgatattactgaatat 1716
Db    874 CTGTTGGCTCCTGGGACCCTGTGCCGCGAGCAGGCCAGGCAGTGTGACCTCCCGGAGTTC 933

Qy   1717 tgtactggagactctggtcagtgcccaccaaatcttcataagcaagacggatatgcatgc 1776
Db    934 TGTACGGGCAAGTCTCCCCACTGCCCTACCAACTTCTACCAGATGGATGGTACCCCCTGT 993

Qy   1777 aatcaaaatcagggccgctgctacaatggcgagtgcaagaccagagacaaccagtgtcag 1836
Db    994 GAGGGCGGCCAGGCCTACTGCTACAACGGCATGTGCCTCACCTACCAGGAGCAGTGCCAG 1053

Qy   1837 tacatctggggaacaaaggctgcagggtctgacaagttctgctatgaaaagctgaataca 1896
Db   1054 CAGCTGTGGGGACCCGGAGCCCGACCTGCCCCTGACCTCTGCTTCGAGAAGGTGAATGTG 1113

Qy   1897 gaaggcac 1904
Db   1114 GCAGGGAC 1121
```

```
RESULT 15
PCT-US93-05640-10
; Sequence 10, Application PC/TUS9305640
;  GENERAL INFORMATION:
;    APPLICANT:  University of Connecticut
;    TITLE OF INVENTION:  Contraceptive Vaccine
;    NUMBER OF SEQUENCES:  46
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  PCT/US93/05640
;      FILING DATE:  19930610
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/897,883
;      FILING DATE:  June 12, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brook, David E.
;      REGISTRATION NUMBER:  22,592
;      REFERENCE/DOCKET NUMBER:  UCT90-01AA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 861-6240
;      TELEFAX:  (617) 861-9540
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1463 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    [redacted]-US93-05640-10


  Query Match              2.9%;  Score 71.8;  DB 5;  Length 1463;
  Best Local Similarity   48.2%;  Pred. No. 2.2e-10;
  Matches  202;  Conservative    0;  Mismatches  217;  Indels    0;  Gaps    0;

Qy   1506 cacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgtggttttcatgt 1565
          | |||   || || |||      ||||||   || || || || || || || ||  || |
Db     27 CCCGGTCTGCGGCAATAACAGGGTGGAACAGGGTGAAGACTGCGACTGCGGATCGCAGGA 86

Qy   1566 ggaatgctatggattatgctgtaagaaatgttccctctccaacggggctcactgcagcga 1625
          ||||||| | |     ||||||  |      | ||  |           |   |   |
Db     87 GGAATGCCAAGACACCTGCTGTGATGCTGCTACCTGCAGGCTGAAAAGTACTTCACGATG 146

Qy   1626 cgggccctgctgtaacaatacctcatgtctttttcagccacgagggtatgaatgccggga 1685
           |  |   |          || |   |||   ||  |  |   |||  | | ||||||  ||
Db    147 TGCTCAAGGGCCCTGTTGTAACCAGTGTGAGTTCAAAACTAAAGGAGAGGTATGCCGAGA 206

Qy   1686 tgctgtgaacgagtgtgatattactgaatattgtactggagactctggtcagtgcccacc 1745
             |   | | |||||||||  |  |||| || || | |  ||    |||||    |||| |
Db    207 GTCCACGGATGAGTGTGATCTCCCTGAGTACTGCAACGGCTCGTCTGGGGCTTGCCAAGA 266

Qy   1746 aaatcttcataagcaagacggatatgcatgcaatcaaaatcagggccgctgctacaatgg 1805
           | | ||  ||       | ||  |   ||||    |  |  || |   ||||   |||||
Db    267 AGACCTCTATGTTATTAATGGGCACAGATGCGCAAATGAAGAGTGGATCTGCATGAATGG 326

Qy   1806 cgagtgcaagaccagagacaaccagtgtcagtacatctggggaacaaaggctgcagggtc 1865
             |||   | |  | |    || || ||  | |  |  || ||| |    | | |
Db    327 GAGGTGTCTGTCTGGGAAGGCACAATGCCAAGAAACATTTGGTACAGAAATGGAAATGGG 386

     1866 tgacaagttctgctatgaaaagctgaatacagaaggcactgagaagggaaactgcggga 1924
            |       ||||| |||| | || |||||  |       |  |||||||| || |
      387 TTCAGTAGACTGCTTTGAACAACTTAATACTAAGAATGACATTACTGGAAACTGTGGTA 445


RESULT  16
US-08-243-542-9
; Sequence 9, Application US/08243542
; Patent No. 5552526
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  :  EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/243,542
;      FILING DATE:
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
;      TELEFAX:  (616) 381-5465
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9278 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY:  human DNA cosmid library
;    FEATURE:
;      NAME/KEY:  exon 1
;      LOCATION:  28..44
;    FEATURE:
;      NAME/KEY:  exon 2
;      LOCATION:  308..374
;    FEATURE:
;      NAME/KEY:  exon 3
;      LOCATION:  909..994
;    FEATURE:
;      NAME/KEY:  exon 4
;      LOCATION:  1081..1156
;    FEATURE:
;      NAME/KEY:  exon 5
;      LOCATION:  1591..1657
;    FEATURE:
;      NAME/KEY:  exon 6
;      LOCATION:  1725..1792
;    FEATURE:
;      NAME/KEY:  exon 7
;      LOCATION:  2182..2256
;    FEATURE:
;      NAME/KEY:  exon 8
;      LOCATION:  2339..2410
;    FEATURE:
;      NAME/KEY:  exon 9
;      LOCATION:  2588..2754
;    FEATURE:
;      NAME/KEY:  exon 10
;      LOCATION:  3248..3332
;    FEATURE:
;      NAME/KEY:  exon 11
;      LOCATION:  3445..3535
;    FEATURE:
;      NAME/KEY:  exon 12
```

```
;      LOCATION:  3645..3696
;    FEATURE:
;      NAME/KEY:  exon 13
;      LOCATION:  4014..4113
;    FEATURE:
;      NAME/KEY:  exon 14
;      LOCATION:  4196..4267
;    FEATURE:
;      NAME/KEY:  exon 15
;      LOCATION:  4386..4478
;    FEATURE:
;      NAME/KEY:  exon 16
;      LOCATION:  4920..5000
;    FEATURE:
;      NAME/KEY:  exon 17
;      LOCATION:  5347..5397
;    FEATURE:
;      NAME/KEY:  exon 18
;      LOCATION:  5501..5564
;    FEATURE:
;      NAME/KEY:  exon 19
;      LOCATION:  5767..5866
;    FEATURE:
;      NAME/KEY:  exon 20
;      LOCATION:  6073..6202
;    FEATURE:
;      NAME/KEY:  exon 21
;      LOCATION:  6300..6468
;    FEATURE:
;      NAME/KEY:  exon 22
;      LOCATION:  6557..6671
;    FEATURE:
;      NAME/KEY:  exon 23
;      LOCATION:  6756..6846
;    FEATURE:
;      NAME/KEY:  exon 24
;      LOCATION:  7829..7846
;    FEATURE:
;      NAME/KEY:  exon 25
;      LOCATION:  8165..9038
US-08-243-542-9


  Query Match             2.7%;  Score 66.6;  DB 1;  Length 9278;
  Best Local Similarity   80.4%;  Pred. No. 1.7e-08;
  Matches   78;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy    2359 ggtcctagtgccaccaatctcataataggctccatcgctggtgccatcctggtagcagct 2418
           ||||| || | ||||||  |||| || ||||||||||||||| ||  ||||||| |||||
Db    6755 GGTCCCAGCGGCACCAACATCATCATTGGCTCCATCGCTGGGGCTGTCCTGGTTGCAGCC 6814

Qy    2419 attgtccttgggggcacaggctggggatttaaaaatg 2455
           || ||||| || |||||| ||||||||||||||  | |
Db    6815 ATCGTCCTGGGCGGCACGGGCTGGGGATTTAAGTAAG 6851


RESULT  17
US-08-477-407-9
; Sequence 9, Application US/08477407
; Patent No. 5631351
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  :  EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/477,407
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/243,542
;      FILING DATE:  13-MAY-1994
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
;      TELEFAX:  (616) 381-5465
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9278 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY:  human DNA cosmid library
;    FEATURE:
;      NAME/KEY:  exon 1
;      LOCATION:  28..44
;    FEATURE:
;      NAME/KEY:  exon 2
;      LOCATION:  308..374
;    FEATURE:
;      NAME/KEY:  exon 3
;      LOCATION:  909..994
;    FEATURE:
;      NAME/KEY:  exon 4
;      LOCATION:  1081..1156
;    FEATURE:
;      NAME/KEY:  exon 5
;      LOCATION:  1591..1657
;    FEATURE:
;      NAME/KEY:  exon 6
;      LOCATION:  1725..1792
;    FEATURE:
;      NAME/KEY:  exon 7
;      LOCATION:  2182..2256
;    FEATURE:
;      NAME/KEY:  exon 8
;      LOCATION:  2339..2410
;    FEATURE:
;      NAME/KEY:  exon 9
;      LOCATION:  2588..2754
;    FEATURE:
;      NAME/KEY:  exon 10
;      LOCATION:  3248..3332
;    FEATURE:
;      NAME/KEY:  exon 11
;      LOCATION:  3445..3535
```

```
;    FEATURE:
;      NAME/KEY:  exon 12
;      LOCATION:  3645..3696
;    FEATURE:
;      NAME/KEY:  exon 13
;      LOCATION:  4014..4113
;    FEATURE:
;      NAME/KEY:  exon 14
;      LOCATION:  4196..4267
;    FEATURE:
;      NAME/KEY:  exon 15
;      LOCATION:  4386..4478
;    FEATURE:
;      NAME/KEY:  exon 16
;      LOCATION:  4920..5000
;    FEATURE:
;      NAME/KEY:  exon 17
;      LOCATION:  5347..5397
;    FEATURE:
;      NAME/KEY:  exon 18
       LOCATION:  5501..5564
     FEATURE:
       NAME/KEY:  exon 19
       LOCATION:  5767..5866
;    FEATURE:
;      NAME/KEY:  exon 20
;      LOCATION:  6073..6202
;    FEATURE:
;      NAME/KEY:  exon 21
;      LOCATION:  6300..6468
;    FEATURE:
;      NAME/KEY:  exon 22
;      LOCATION:  6557..6671
;    FEATURE:
;      NAME/KEY:  exon 23
;      LOCATION:  6756..6846
;    FEATURE:
;      NAME/KEY:  exon 24
;      LOCATION:  7829..7846
;    FEATURE:
;      NAME/KEY:  exon 25
;      LOCATION:  8165..9038
US-08-477-407-9


  Query Match             2.7%;  Score 66.6;  DB 1;  Length 9278;
  Best Local Similarity   80.4%;  Pred. No. 1.7e-08;
  Matches   78;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy    2359 ggtcctagtgccaccaatctcataataggctccatcgctggtgccatcctggtagcagct 2418
           ||||| || | ||||||  |||| || ||||||||||||||| ||  ||||||| |||||
       6755 GGTCCCAGCGGCACCAACATCATCATTGGCTCCATCGCTGGGGCTGTCCTGGTTGCAGCC 6814

       2419 attgtccttgggggcacaggctggggatttaaaaatg 2455
            || ||||| || ||||| ||||||||||||||  | |
       6815 ATCGTCCTGGGCGGCACGGGCTGGGGATTTAAGTAAG 6851


RESULT  18
US-08-484-355-9
; Sequence 9, Application US/08484355
; Patent No. 5705341
;  GENERAL INFORMATION:
;    APPLICANT:  NAKAMURA, YUSUKE
;    APPLICANT:  :  EMI, MITSURU
;    TITLE OF INVENTION:  MDC PROTEINS AND DNAs
;    TITLE OF INVENTION:  ENCODING THE SAME
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  FLYNN, THIEL, BOUTELL & TANIS P.C.
;      STREET:  2026 Rambling Road
;      CITY:  Kalamazoo
;      STATE:  Michigan
;      COUNTRY:  USA
;      ZIP:  49008-1699
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inches, 1.44 Mb storage
;      COMPUTER:  IBM PC/XT/AT Compatible
;      OPERATING SYSTEM:  MS-DOS 5.0
;      SOFTWARE:  WordPerfect 5.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,355
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/243,542
;      FILING DATE:  13-MAY-1994
;      APPLICATION NUMBER:  JP 5-136602
;      FILING DATE:  14 MAY 1993
;      APPLICATION NUMBER:  JP 5-257455
;      FILING DATE:  22 SEPTEMBER 1993
;      APPLICATION NUMBER:  JP 6-49904
;      FILING DATE:  23 FEBRUARY 1994
;      APPLICATION NUMBER:  JP 6-73328
;      FILING DATE:  12 APRIL 1994
;      APPLICATION NUMBER:  JP 6-84470
;      FILING DATE:  22 APRIL 1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Terryence F. Chapman
;      REGISTRATION NUMBER:  32 549
;      REFERENCE/DOCKET NUMBER:  Furuya Case 1313
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (616) 381-1156
;      TELEFAX:  (616) 381-5465
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9278 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Homo sapiens
;    IMMEDIATE SOURCE:
;      LIBRARY:  human DNA cosmid library
;    FEATURE:
;      NAME/KEY:  exon 1
;      LOCATION:  28..44
;    FEATURE:
;      NAME/KEY:  exon 2
;      LOCATION:  308..374
;    FEATURE:
;      NAME/KEY:  exon 3
;      LOCATION:  909..994
;    FEATURE:
;      NAME/KEY:  exon 4
;      LOCATION:  1081..1156
;    FEATURE:
;      NAME/KEY:  exon 5
;      LOCATION:  1591..1657
;    FEATURE:
;      NAME/KEY:  exon 6
;      LOCATION:  1725..1792
;    FEATURE:
;      NAME/KEY:  exon 7
;      LOCATION:  2182..2256
;    FEATURE:
;      NAME/KEY:  exon 8
;      LOCATION:  2339..2410
;    FEATURE:
;      NAME/KEY:  exon 9
;      LOCATION:  2588..2754
;    FEATURE:
;      NAME/KEY:  exon 10
;      LOCATION:  3248..3332
;    FEATURE:
```

```
;       NAME/KEY:  exon 11
;       LOCATION:  3445..3535
;     FEATURE:
;       NAME/KEY:  exon 12
;       LOCATION:  3645..3696
;     FEATURE:
;       NAME/KEY:  exon 13
;       LOCATION:  4014..4113
;     FEATURE:
;       NAME/KEY:  exon 14
;       LOCATION:  4196..4267
;     FEATURE:
;       NAME/KEY:  exon 15
;       LOCATION:  4386..4478
      FEATURE:
        NAME/KEY:  exon 16
        LOCATION:  4920..5000
      FEATURE:
;       NAME/KEY:  exon 17
;       LOCATION:  5347..5397
;     FEATURE:
;       NAME/KEY:  exon 18
;       LOCATION:  5501..5564
;     FEATURE:
;       NAME/KEY:  exon 19
;       LOCATION:  5767..5866
;     FEATURE:
;       NAME/KEY:  exon 20
;       LOCATION:  6073..6202
;     FEATURE:
;       NAME/KEY:  exon 21
;       LOCATION:  6300..6468
;     FEATURE:
;       NAME/KEY:  exon 22
;       LOCATION:  6557..6671
;     FEATURE:
;       NAME/KEY:  exon 23
;       LOCATION:  6756..6846
;     FEATURE:
;       NAME/KEY:  exon 24
;       LOCATION:  7829..7846
;     FEATURE:
;       NAME/KEY:  exon 25
;       LOCATION:  8165..9038
US-08-484-355-9


 Query Match             2.7%;  Score 66.6;  DB 1;  Length 9278;
 Best Local Similarity   80.4%;  Pred. No. 1.7e-08;
 Matches   78;  Conservative    0;  Mismatches   19;  Indels    0;  Gaps    0;

Qy    2359 ggtcctagtgccaccaatctcataataggctccatcgctggtgccatcctggtagcagct 2418
           ||||| || | ||||||  |||| || ||||||||||||||| ||  ||||||| |||||
Db    6755 GGTCCCAGCGGCACCAACATCATCATTGGCTCCATCGCTGGGGCTGTCCTGGTTGCAGCC 6814

Qy    2419 attgtccttgggggcacaggctggggatttaaaaatg 2455
           || ||||| || ||||| ||||||||||||||  | |
Db    6815 ATCGTCCTGGGCGGCACGGGCTGGGGATTTAAGTAAG 6851


RESULT  19
US-08-264-101-1
; Sequence 1, Application US/08264101
; Patent No. 5693496
;  GENERAL INFORMATION:
;     APPLICANT:  ALVES, KENNETH
;     APPLICANT:  GUPTA, SUNIL K.
;     APPLICANT:  HOLLIS, GREGORY F.
;     TITLE OF INVENTION:  CONTRACEPTIVE VACCINE
;     NUMBER OF SEQUENCES:  4
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  MARY A. APPOLLINA
;       STREET:  P.O. BOX 2000, 126 E. LINCOLN AVENUE
;       CITY:  RAHWAY
;       STATE:  NJ
;       COUNTRY:  USA
;       ZIP:  07065
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/264,101
;       FILING DATE:
;       CLASSIFICATION:  424
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  APPOLLINA, MARY A.
;       REGISTRATION NUMBER:  34,087
;       REFERENCE/DOCKET NUMBER:  19244
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (908)594-3462
;       TELEFAX:  (908)594-4720
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  2373 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-264-101-1


  Query Match             2.6%;  Score 64;  DB 1;  Length 2373;
  Best Local Similarity   48.4%;  Pred. No. 4.3e-08;
  Matches  243;  Conservative    0;  Mismatches  250;  Indels    9;  Gaps    2;

Qy    1502 agcccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgtggttttc 1561
           |||   | |  ||||| |||| | |  ||||||| || |||||||||||| |||||   |
Db     896 AGCAAGCAGTGTGTGGTAATGCAAAGCTGGAAGCAGGAGAGGAGTGTGACTGTGGGACTG 955

Qy    1562 atgtggaa--tgctatggattatgctgtaagaaatgttccctctccaacggggctcactg 1619
           |   |||   |||  |   |   |    | |   || |    |  ||   |  | |  |
Db     956 AACAGGATTGTGCCCTTATTGGAGAAACATGCTGTGATATTGCCACATGTAGATTTAAAG 1015

Qy    1620 cagcgacgggccctgctgtaacaata-------cctcatgtctttttcagccacgagggt 1672
           | |   |   |  ||||| |  |  |            ||||| |||  | ||  ||
Db    1016 CCGGTTCAAACTGTGCTGAAGGACCATGCTGCGAAAACTGTCTATTTATGTCAAAAGAAA 1075

Qy    1673 atgaatgccgggatgctgtgaacgagtgtgatattactgaatattgtactggagactctg 1732
             ||  ||  | |  |  | || || ||  |  |||||||||| | ||||   ||||
Db    1076 GAATGTGTAGGCCTTCCTTTGAAGAATGCGACCTCCCTGAATATTGCAATGGATCATCTG 1135

Qy    1733 gtcagtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatcagggcc 1792
             ||||||  ||| |   ||   ||    ||  || | ||    |  |||||  |
Db    1136 CATCATGCCCAGAAAACCACTATGTTCAGACTGGGCATCCGTGTGGACTGAATCAATGGA 1195

Qy    1793 gctgctacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctggggaacaa 1852
            |||     |||| |  || | ||   | || || || |||    |||  |  || | |
Db    1196 TCTGTATAGATGGAGTTTGTATGAGTGGGGATAAACAATGTACAGACACATTTGGCAAAG 1255

Qy    1853 aggctgcagggtctgacaagttctgctatgaaaagctgaatacagaaggcactgagaagg 1912
           | | |                  || |||    | || ||| || |      ||     |
Db    1256 AAGTAGAGTTTGGCCCTTCAGAATGTTATTCTCACCTTAATTCAAAGACTGATGTATCTG 1315

Qy    1913 gaaactgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtg 1972
           ||||||| || |   ||      | |  | |||||        |  || ||   || |
Db    1316 GAAACTGTGGTATAAGTGATTCAGGATACACACAGTGTGAAGCTGACAATCTGCAGTGCG 1375

Qy    1973 gattcttactctgtaccaatct 1994
           ||   ||| | ||||   || |
Db    1376 GAAAATTAATATGTAAATATGT 1397


RESULT  20
US-08-765-243-1
```

```
; Sequence 1, Application US/08765243
; Patent No. 5935578
;  GENERAL INFORMATION:
;    APPLICANT:  ALVES, KENNETH
;    APPLICANT:  GUPTA, SUNIL K.
;    APPLICANT:  HOLLIS, GREGORY F.
;    TITLE OF INVENTION:  CONTRACEPTIVE VACCINE
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MARY A. APPOLLINA
;      STREET:  P.O. BOX 2000, 126 E. LINCOLN AVENUE
;      CITY:  RAHWAY
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07065
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
     CURRENT APPLICATION DATA:
       APPLICATION NUMBER:  US/08/765,243
       FILING DATE:
       CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  APPOLLINA, MARY A
;      REGISTRATION NUMBER:  34,087
;      REFERENCE/DOCKET NUMBER:  19244Y
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (908)594-3462
;      TELEFAX:  (908)594-4720
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2373 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-765-243-1


  Query Match             2.6%;  Score 64;  DB 2;  Length 2373;
  Best Local Similarity   48.4%;  Pred. No. 4.3e-08;
  Matches  243;  Conservative    0;  Mismatches  250;  Indels    9;  Gaps    2;

Qy   1502 agcccacggaatgtggaaatggatacgtggaagctggggaggagtgtgattgtggttttc 1561
          |||   | |  ||||| |||| | |  ||||||| || ||||||||||| |||||   |
Db    896 AGCAAGCAGTGTGTGGTAATGCAAAGCTGGAAGCAGGAGAGGAGTGTGACTGTGGGACTG 955

Qy   1562 atgtggaa--tgctatggattatgctgtaagaaatgttccctctccaacggggctcactg 1619
          |   |||   |||  |   |   |    | |   || |  .  |  ||    |  | |  |
     ?     956 AACAGGATTGTGCCCTTATTGGAGAAACATGCTGTGATATTGCCACATGTAGATTTAAAG 1015

     ?    1620 cagcgacgggccctgctgtaacaata-------cctcatgtctttttcagccacgagggt 1672
          | |  |   |  |||||  |  |  |            |||||  |||  |  ||  ||
Db   1016 CCGGTTCAAACTGTGCTGAAGGACCATGCTGCGAAAACTGTCTATTTATGTCAAAAGAAA 1075

Qy   1673 atgaatgccgggatgctgtgaacgagtgtgatattactgaatattgtactggagactctg 1732
          ||  ||  | |  |  | || || ||  |  ||||||||||  | ||||   ||||
Db   1076 GAATGTGTAGGCCTTCCTTTGAAGAATGCGACCTCCCTGAATATTGCAATGGATCATCTG 1135

Qy   1733 gtcagtgcccaccaaatcttcataagcaagacggatatgcatgcaatcaaaatcagggcc 1792
            ||||||  ||| |   ||   ||    ||  || | ||    |  |||||  |
Db   1136 CATCATGCCCAGAAAACCACTATGTTCAGACTGGGCATCCGTGTGGACTGAATCAATGGA 1195

Qy   1793 gctgctacaatggcgagtgcaagaccagagacaaccagtgtcagtacatctggggaacaa 1852
           |||     |||| |  || | ||   | || || || |||    |||  |  || | |
Db   1196 TCTGTATAGATGGAGTTTGTATGAGTGGGGATAAACAATGTACAGACACATTTGGCAAAG 1255

Qy   1853 aggctgcagggtctgacaagttctgctatgaaaagctgaatacagaaggcactgagaagg 1912
          | |  |                 || |||    | || ||| || |      ||     |
Db   1256 AAGTAGAGTTTGGCCCTTCAGAATGTTATTCTCACCTTAATTCAAAGACTGATGTATCTG 1315

Qy   1913 gaaactgcgggaaggatggagaccggtggattcagtgcagcaaacatgatgtgttctgtg 1972
          ||||||| || |    ||       | |  |  |||||        |  || ||   || |
Db   1316 GAAACTGTGGTATAAGTGATTCAGGATACACACAGTGTGAAGCTGACAATCTGCAGTGCG 1375

Qy   1973 gattcttactctgtaccaatct 1994
          ||   ||| | ||||   || |
Db   1376 GAAAATTAATATGTAAATATGT 1397


Search completed: August 15, 2001, 21:53:11
Job time: 7579 sec
```

THIS PAGE BLANK (USPTO)